

Optimal Resource Allocation and Quality of Service Prediction in Cloud

Priya Baldoss^{1,2,*} and Gnanasekaran Thangavel³

¹Information and Communication Engineering Department, Anna University, Chennai, India

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

³RMK Engineering College, Chennai, India

*Corresponding Author: Priya Baldoss. Email: priyaannauniversity.phd@gmail.com

Received: 17 August 2020; Accepted: 21 October 2020

Abstract: In the present scenario, cloud computing service provides on-request access to a collection of resources available in remote system that can be shared by numerous clients. Resources are in self-administration; consequently, clients can adjust their usage according to their requirements. Resource usage is estimated and clients can pay according to their utilization. In literature, the existing method describes the usage of various hardware assets. Quality of Service (QoS) needs to be considered for ascertaining the schedule and the access of resources. Adhering with the security arrangement, any additional code is forbidden to ensure the usage of resources complying with QoS. Thus, all monitoring must be done from the hypervisor. To overcome the issues, Robust Resource Allocation and Utilization (RRAU) approach is developed for optimizing the management of its cloud resources. The work hosts a numerous virtual assets which could be expected under the circumstances and it enforces a controlled degree of QoS. The asset assignment calculation is heuristic, which is based on experimental evaluations, RRAU approach with J48 prediction model reduces Job Completion Time (JCT) by 4.75 s, Make Span (MS) 6.25, and Monetary Cost (MC) 4.25 for 15, 25, 35 and 45 resources are compared to the conventional methodologies in cloud environment.

Keywords: Cloud computing; resource utilization; robust resource allocation and utilization (RRAU) approach; job completion time; quality of services; monetary cost; make span

1 Introduction

Cloud computing offers an on-request access to a pool of processing and a capacity asset available over the system that is shared by numerous clients. Assets are in self-administration; so that, the clients can adjust their usage as per their requirements. Asset usage is estimated and clients pay according to their utilization. Thus, clients are financially incentivized to discharge the assets, which are not required and they can assist different clients. The asset allocation algorithm is a heuristic approach. The nature of choices relies upon how well the heuristic suits the remaining workload.



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Previous works have described the organization of Virtual Machine (VM) dismissing the use of volumes, previews and security gatherings. Since these virtual assets are utilized in relationship with VMs, there is a lot to be discovered from the characterization of co-organizations. For example, the portrayal of the related usage of volumes and depictions are required to improve the capacity administration. Furthermore, a workload characterization in the previous study, concentrates on hyper scale stages facilitating a large number of VMs every month. Concentrating on VMs, past works reveal the usage of hardware assets CPU, RAM and IO. According to the observation, no analysis is carried on the QoS as a whole. To characterize the usage of hardware assets and QoS, a protection approach becomes essential, and it has to be adopted by the hypervisor. The past perceptions led the users to put forth four research questionnaires: (1) How do cloud end users deploy and communicate with virtual assets? (2) By which way, such contrasts can be utilized to enhance resource allocations? (3) Which QoS measurements are noticeable from the hypervisor and (4) what elements impact them? In this scenario, asset assignment can be characterized as the mapping between the provider's hardware and the clients' virtual assets. Subsequently, the provider looks to limit fragmentation, which is the presence of accessible assets conveyed over the infrastructure, however in sums, too little to ever be allocated. Since servers comprise of CPU, RAM, and other different assets, such as GPUs, the fragmentation of an asset mostly relies upon the usage of others, where the provider needs to limit administration costs. Over responsibility is characterized as the assignment of an unbreakable asset, for example, a CPU center, to several clients.

Robust Resource Allocation and Utilization (RRAU) approach proposes to optimize the management of its cloud resources. This work hosts a number of virtual assets that implements a controlled degree of QoS. The asset allocation algorithm is a heuristic approach. The nature of choices relies upon how well the heuristic suits the workload. Relocations are time consuming since the server continues to perform without expecting the clients to stop and restart their VMs. Unluckily, VM migration requires some energy, devours the assets, and degrades the QoS of VMs and in some cases, servers fail with heterogeneous attributes. RRAU approach predicts the runtime of VMs that depends on metadata at start-up. It shows that the usages of labels, which are freely-typed pieces of content, are used to portray VMs which improves altogether characterization prediction results. The paper contribution is expressed as:

- To develop RRAU approach for optimizing the management of its cloud resources and implement a controlled degree of QoS
- To address the problem of improvising resource optimization by characterizing the workload for identifying the opportunities regarding resource management
- To predict the runtime of VMs dependent on metadata, accessible at start-up for finding the solutions to balance the workload
- To reduce Job Completion Time (JCT), Make Span (MS), and Monetary Cost (MC) for resources this approach is compared to the conventional methodologies in cloud environment

The remaining paper is framed as follows: Section 2 explains about the recent work of optimal resource utilization in cloud with prediction mechanisms. Section 3 elaborates the proposed methodology, executions steps, and algorithms details along with its features and implementation steps. Section 4 discusses experimental setup, input details, evaluation metrics and comparative result analysis. Section 5 concludes the overall research work with the suggestions for future plan.

2 Literature Work

Moreno-Vozmediano et al. [1] discussed and assessed a prescient auto-scaling mechanism dependent on machine learning methods for time forecasting and queuing theory. The system predicts the preparing workload of a conveyed server and the appraisals with a suitable number of assets that must be provisioned so as to streamline the service response time and satisfy the Service Level Agreement (SLA) shrunk by the client. Haouari et al. [2] addressed a prediction driven asset allocation structure, to augment the Quality of Experience (QoE) of viewers and limit the asset distribution cost. It executed a machine learning model to anticipate the viewer's number close to each geo-distributed cloud environment. The technique formulates an advancement issue to proactively allocate assets at the viewer's proximity. Bashir et al. [3] tended to asset assignment issue in 5G systems, in the Cloud Radio Access Networks (C-RAN). The Radio Access Network (RAN) systems include various system topologies that are confined dependent on the range groups and they ought to be upgraded with various access technologies in the organization of 5G arrangement. C-RAN is one of the ideal systems which are used to join all the accessible spectral bands. Hu et al. [4] proposed a prediction model to estimate the changing JCT of a single Spark job as described. With the support of the prediction method, the algorithm balances the resource allocation of multiple Spark jobs, aiming to minimize the average JCT in multiple-job cases. Thang et al. [5] examined the issue of dependable asset provisioning in joint edge-cloud conditions, studied methodologies and strategies that can be utilized to improve the reliability of the distributed applications in various heterogeneous network situations. Because of the multifaceted nature of the issue, specific accentuation is set on answers for the portrayal, the executives, and the control of complex distributed applications using machine learning methods.

Afrin et al. [6] dealt with a synchronous enhancement of Make Span, vitality utilization and cost while assigning assets for the undertakings of an automated work process. The method developed an Edge Cloud based multi-robot framework to overcome the confinements of remote cloud based framework in converting delay sensitive data. Zafari et al. [7] explained an asset sharing structure that permits various ESPs to ideally use their assets and improve the fulfillment level of uses subject to imperatives, for example, communication cost for sharing assets across ESPs. The system thinks about various ESPs that have their own destinations for using their assets, resulting in a multi-objective optimization issue. Mandal et al. [8] annealing-based optimized load balancing adjusting by including VM migration arrangement starting with one host, then on to the other linear regression-based prediction policy for cutting edge asset usage is simulated. Inactive load balancing approaches are avoided to guarantee QoS, without missing the deadline by the appropriation of dynamic workload evenly. Wajahat [9] elaborated a model-driven solution for reallocating computational resources for the existing networking services. The method focuses on addressing the resource management challenges faced by cloud providers when delivering cloud services to their tenants or clients. Elgendy et al. [10] described the restrictions of such devices. To start with, the computation and radio assets are mutually considered for multiuser situations to ensure the productive use of the shared assets. Moreover, an Advanced Encryption Standard (AES) cryptographic strategy is acquainted as a security layer with the protected sensitive data from cyber-attacks.

Yu et al. [11] focused on asset distribution for TV multimedia service assistance in the 5G wireless cloud arrangement (C-RAN) situation, which can assist unicast services for cellular clients and multicast administrations for broadcast services all the while. It structured the relative reduction assets distribution architecture depending on the idea of a self-organizing system. The management architecture first builds the capacities and procedures of the relating independent

asset of the management. Podolskiy et al. [12] investigated an answer for the self-versatile issue of vertical flexibility for co-located containerized applications. It determines the limits that meet SLAs and also the asset utilization through an integration of advancement and the confined brute-force search. Sniezynski et al. [13] tended to have a reservation plan adjustment framework dependent on machine learning. With regard to cloud auto-scaling, a significant issue is the capacity to characterize and utilize an asset reservation plan, which empowers asset scheduling. It permits the updating of a reservation plan, initially arranged by an admin. Chen et al. [14] explained a benefit of actor-critic base Reinforcement Learning (RL) system for asset distribution in cloud datacenter. It parameterizes the planning (distributing assets) and it selects constant activities (scheduling tasks) in light of the scores (assessing activities). Gao et al. [15] contemplated the issue of allocating Virtual Machine (VM) assets in geo-distributed ECNs to mobile clients by utilizing the auction hypothesis. It thinks about mobile clients and ECNs as the purchasers and dealers of the VM asset auction, individually.

Aziz et al. [16] evaluated the different works that focused on asset management and information processing in Bigdata platforms. Moreover, it produces adaptability examination utilizing Spark. The technique evaluates the speedup and the handling time. It also finds a specific number of hubs in cluster. Tchernykh et al. [17] explained the job of uncertainty in the asset and administration provisioning, protection, within the sight of the dangers of classification, uprightness, and accessibility. The method reviewed the sources of uncertainty, and essential methodologies for planning under sources of uncertainty. Zeng et al. [18] elaborated energy-efficient methodologies for transmission bandwidth distribution and scheduling. They adjust to devices' channel states and estimation limits in order to lessen their total energy utilization while justifying learning execution. Wei et al. [19] expounded cloud application auto-scaling approach dependent on Q-learning technique to help Software as a Service (SaaS) suppliers to settle on ideal asset portion choices in a dynamic and stochastic cloud condition. It considered diverse VM pricing mechanisms in model, including on-demand and reserved pattern. Arunarani et al. [20] discussed a comprehensive study of task scheduling systems and the related metrics suitable for cloud which are examined with the identification of different issues with the scheduling methodologies and the impediments to survive.

3 Proposed Methodology

The section presents a model to predict the runtime of VMs dependent on metadata accessible at start-up. It shows that, the usage of labels which are freely-typed pieces of content portrays VMs and it improves the characterization prediction results. Fig. 1 displays the workflow of the proposed algorithm with systematic representations.

3.1 Workload Models

VM placement is an expansion issue, which holds the proposed framework objective: reduction of the asset wastage and estimation of SLA. The solution for the candidate's issue is analyzed, and a target work is devised, which accepts the asset usage as its input, and it exhibits the VMs.

3.2 Static Resource Usage Model

The imprint of a VM is displayed by a static use of assets. The asset utilization is multi-dimensional and it contributes to a small amount of a server's assets CPU, RAM, disk, and system bandwidth. The framework proposes a maximum vector to offer solutions to all datacenter

assets. Asset usage needs more memory to be encoded. Execution is viewed as adequate, as long as the absolute use of an asset is not exactly an advantage. The model is candid until the asset use changes, or the existing VMs stop, or the new VMs start. Whenever changes happen, the provider must re-optimize asset distribution via costly migrations.

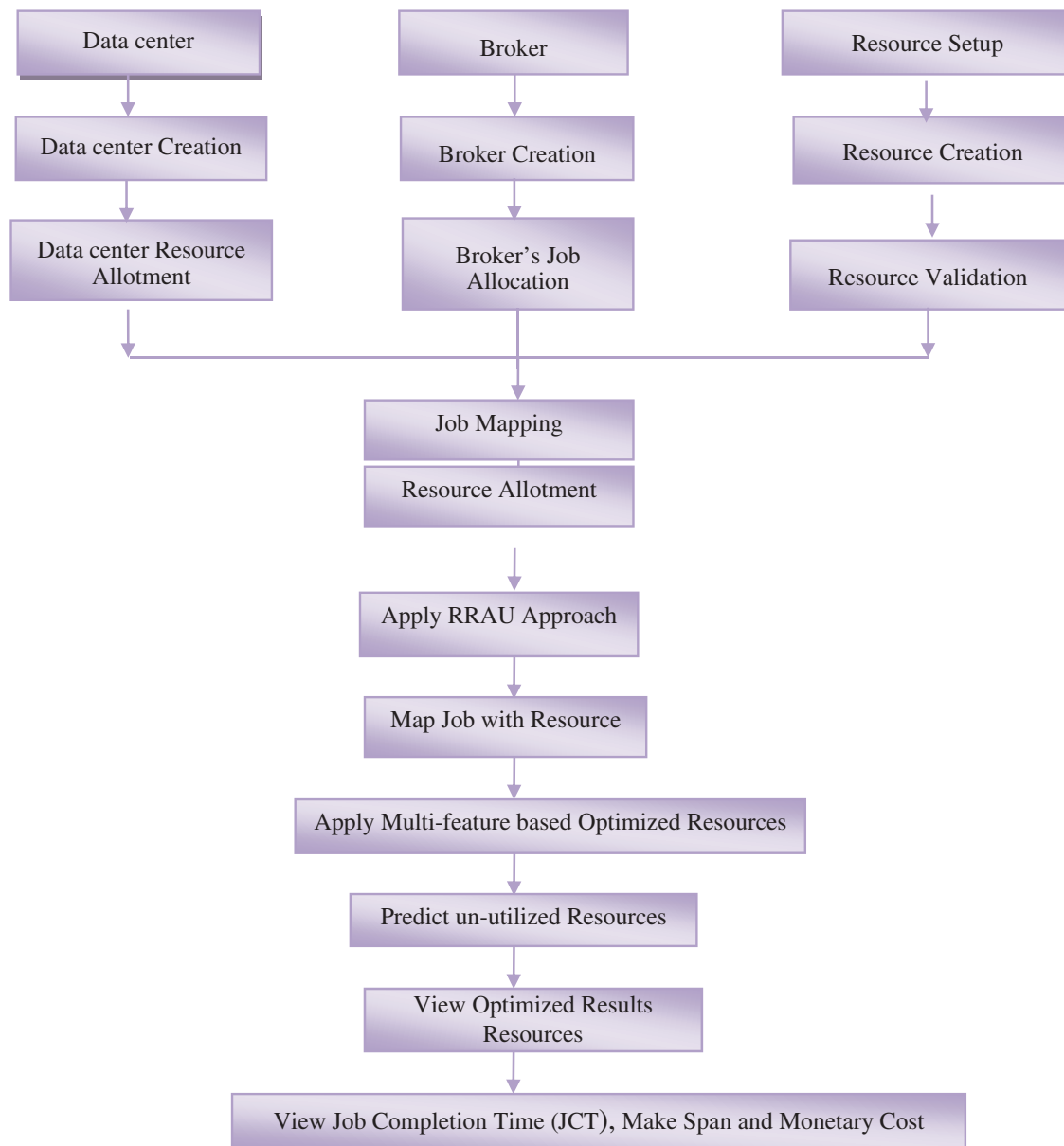


Figure 1: Workflow of RRAU approach

3.3 Dynamic Resource Usage Model

The dynamic usage model concentrates on long-term improvements, as it catches changes in the asset use of VMs. The impression of a VM is displayed by a time series. The exhibition of a

VM is evaluated from the relationship of the asset used by the VM, and the total use of neighbor VMs. The model assumes that VMs run uncertainly, and that asset usage is occasional.

3.4 Clairvoyant Resource Usage Model

A clairvoyant model accepts the provider knowing both the asset usage and the future plan of VM executive's demands. A completely designed clairvoyant model is considered, where the provider is aware of the planning of VM start and stop demands, under the assumption that VMs start at the same time and their runtime is well-known. It considers the state when the runtime of VMs is known, however not at the initial time. Asset utilization model has three natural impairments. It separates client confronting, latency-sensitive applications such as web servers or database servers, batch applications and the disconnected information analytics. They have diverse QoS prerequisites: impedances are endured for latency-sensitive applications, yet inactivity latency-sensitive applications have a tough cutoff time. The framework presents models that hook heterogeneous execution profiles.

3.5 Multi-Feature Optimized Resource Usage Model

The minimization of resource wastage and the execution of SLAs are conflicting goals. The main view is to regard one goal as a requirement, and it assesses the arrangements dependent on the subsequent target. The algorithm minimizes the quantity of servers under the limitation that, the likelihood of server overload is beneath an edge. The subsequent view is to join the assessment of different goals with a similar capacity. The objective function joins the vitality effectiveness and the execution goals. Considering all the above factors, it leads to the difficulty to decide on the ideal weighting.

3.6 RRAU Approach

RRAU approach is designed to predict the runtime of VMs dependent on metadata accessible at start-up. It shows that the usage of labels which are freely-typed pieces of content portrays VMs, and altogether it improves characterization prediction results. The method evaluates and dissects the sensitivity of a VM placement algorithm from the related work which requires forecasts of VM runtimes to optimize asset and time. The proposed algorithm is expressed to estimate under the inference of perfect classifications. The method decides the necessary forecast precision, investigates the sensitivity of the proposed algorithm as for the classification error. The framework deploys a different degree of classification error and it assesses the asset usage of servers against any-fit and best-fit, well-knowledge VM placement algorithms oblivious of the runtime. The RRAU approach delivers a best-fit by considering the runtime of VMs, where the VM runtime is known when the job request is made. The technique looks to co-locate VMs that will stop in parallel. The framework of this methodology is a more optimal asset which is more effective than Best Fit, since servers can execute an outstanding task at hand in less time than Best Fit.

Cloud datacenters offer a huge pool of assets on-request which eases the scope of VM arrangement on the web. Machine learning estimates the structure of proposed algorithms which figures out how explicit programming can be utilized to accelerate the inquiry of the ideal VM placement. The proposed technique works for changing in asset usage and it takes expert dynamic choices. Machine learning is utilized to foresee upcoming asset usage dependent on past perceptions overloaded and to migrate VMs out of the servers which are to be overloaded. Machine learning can also be utilized to approximate the evacuation profile of VMs, which permits processing a VM placement configuration in a faster process than a precise exhibition model. RRAU

approach is sufficiently versatile to present mechanisms made out of thousands of assets and it makes conceivable to present to both physical and virtual assets using cloud explicit ideas, for example, the infrastructure elasticity. The method decreases the bandwidth issues and it keeps up job arrival process and System Queue. Here, the method effectively assesses the cloud datacenter execution and response time. This model is appropriate for minimal and maximal infrastructure. This model assists to control the client and cloud datacenter locally as well globally. The proposed technique assesses the impacts of various assets using the board management technique on the cloud datacenter working and to foresee the relating costs/benefits. The proposed procedure makes thousands of assets adaptable to represent various arrangements and cloud-specific techniques. RRAU approach is designed to have an effective framework to manage the local and the global cloud datacenter resource allocation and the utilization for optimal utilizations from different regions. The pseudo code of RRAU approach is as follows:

Input: Datacenter D, Virtual Machine VM, Broker B, Job Allocation JA, Resource Allotment RA, Resource Utilization RU

Output: Display Job Completion Time (JCT), MakeSpan (MS), and Monetary Cost (MC)

Procedure:

Create Datacenter;

DC Allotment done with a number of machines, Host id, number of PEs, RAM size and Bandwidth;

Create Broker;

Broker allocation is done with the number of B and job mapping;

Create Resource;

Resource Allotment performed with VM id, Broker id, Memory, RAM and Bandwidth;

Process job mapping;

Perform the resource allotment;

Apply Robust Resource Allocation and Utilization (RRAU) approach

Mapped specific job with resources;

Apply Multi-feature Optimized Resource Usage Model with resource;

If

Predict the utilized resource;

View un-utilized resources

Else

Re-allocate un-utilized resources;

Perform job execution;

Show the optimized resource outcome;

Visualize Job Completion Time (JCT), Make Span, and Monetary Cost (MC)

Pseudo code: RRAU approach

4 Results and Discussions

4.1 Deployment Setup

The experiment is developed on Intel core i6 processor, 8 GB RAM and 500 GB Memory with Windows 7 operating systems. The programming language used is Java with JDK 1.8, Net Beans 8.0.2 with CloudSim library to evaluate the performance of the proposed techniques.

4.1.1 Input Configurations

The experimental setup input configurations are distributed to execute the experiment for evaluating the efficiency of the proposed methodologies. The input parameters are explained in [Tab. 1](#).

Table 1: Cloud experimental details

Parameters	Value
Number of jobs	50–1000
Virtual machine	15
Number of workflow	35–65
Number of resources	15–45
Processing of virtual resources	10000–30000 MIPS
Tasks data size	5000–10000 MI
Monetary cost of virtual resources	0.50–0.90 Cents
Allowable task completion time	4000–5000 ms
Total budget	150–200 Cents
Data dependency threshold	1200–1500 ms
RAM	512–2048
Bandwidth	1000 MBPS
DC VM	Xen
No. of process machine	4
No. of running deployment	60
DC VM policy	Time shared
DC OS	Linux
VM memory	1–10 GB
DC architecture	X86

4.2 Simulation Results

This section presents the experimental data, the results and the result analysis. CloudSim is utilized for checking the presentation of the improved asset usage with algorithmic principle. CloudSim is an extensible simulation toolkit that enables design and simulation of the Cloud computing environments and application provisioning. The CloudSim toolbox offers the model and the features of Cloud components, for example, datacenter, Virtual Machines (VMs) and asset provisioning policies. RRAU approach has limited amount of VMs and variable amount of cloudlets. The proposed method describes the mathematical expression of Job Execution Time (JET), MS and MC which evaluates the efficiency of the proposed techniques.

4.2.1 Job Execution Time (JET)

JET is an average value for every user but it is not standardized according to the work volume. In resource allocation, the JET executes the whole application, which is demanded by users from various cloud brokers. The execution time for scheduling jobs and optimization burdens the VM. It shows the speed with which cloud user applications are answered. $JET_{i,j}$ is the cross product of work time and price on asset j which is expressed in Eq. (1)

$$JET_{ip} = \left(TS_{ip} + \frac{taskload_t}{Ability_p} \right) \times \left(price_p \times \frac{taskload_t}{Ability_p} \right) \quad (1)$$

Only after its parent tasks are executed by the partial ordering relationship of task that begins for each job t_i, j relates to the moment the resource starts and the last time I run on resource j . When his parent's task has been completed and the asset j is free. Workload I refers to job capacity load j , cost j is computing capacity and cost.

4.2.2 Make Span (MS)

MS is calculated by a virtual machine as a total time utilization to complete the entire assigned job as a fixed deadline. Make Span selects the resources for the complete machine (virtual machine, RAM, bandwidth, and memory) which runs after completing all jobs execution. Make Span is expressed in Eqs. (2) and (3),

$$MakeSpan = \max \{ MS_{rj} \} \quad (2)$$

$$MS_{rj} = \sum_{T_i \in \theta_{rj}} EET_{T_i r_j} \quad (3)$$

4.2.3 Monetary Costs (MC)

MC assesses all the expenses for server execution, and client demand appearance to P2P crossover cloud datacenter handling cost. The method estimate is based on a number of jobs required to use the asset for performing task demand response process in the cloud. The MC is determined in condition (4),

$$MC = \sum_{j=1}^n JET \times UCost_j \quad (4)$$

where, n is all out volume of task, which is demanded by cloud clients and JET is job execution time.

Tab. 2 shows JET, MS, and MC for 35, 45, 55 and 65 tasks with conventional methods. The proposed method is evaluated with NSGA-II (Non-dominated Sorting Genetic Algorithm-II) [06], Multi-Objective Particle Swarm Optimization (MOPSO) [6], Strength Pareto Evolutionary Algorithm II (SPEA2) [06] and the Pareto Archived Evolution Strategy (PAES) [6] conventional methods. Multi-Objective Particle Swarm Optimization (MOPSO) Algorithm [6] is described for minimizing continuous functions, where the implementation is bearable, computationally cheap and compressed. The algorithm initially performs the mutation on the entire population, after which it quickly diminishes its inclusion over the long run. The method assists on behalf of preventing premature convergence due to existing nearby Pareto fronts. But, the method is unable to find a different set of solutions and it is converging nearer in order to approach the genuine pareto-optimal set. RRAU approach optimizes the administration of its cloud assets. The work

has the option to have many virtual assets that could be expected under the circumstances and it enforces a controlled degree of QoS. The system illustrates that the methodology is a progressively optimal asset which is more effective and the servers can execute a given workload in less time. The system predicts the runtime of VMs dependent on metadata accessible at start-up. RRAU approach is incorporated with the J48 classifier to decrease classification error and it enhances resource utilization accuracy to deliver the most optimal resource allocation and utilization. The classifier reflects a prediction procedure for predicting categorical data according to their characteristics. It is also efficient in processing large amounts of resource information and it is therefore often used in the implementation of resource allocation. RRAU approach and J48 classifier reduces JCT 4.75 s, MS 12.5, and MC 4.25 for 15, 25, 35 and 45 resources in cloud environment. Based on tabular results, it can be stated that the proposed method performs better when compared to the conventional methods.

Table 2: JCT, MS, and MC for 35, 45, 55 and 65 tasks

Existing algorithms	35			45			55			65		
	JCT	MS	MC	JCT	MS	MC	JCT	MS	MC	JCT	MS	MC
NSGA-II	10.10	65	37	19.33	75	47	27.13	107	54	37.13	111	66
MOPSO	7.93	58	35	14.07	68	42	21.31	88	47	31.31	96	64
SPEA2	8.42	60	40	15.70	85	50	23.36	92	56	33.36	116	68
PAES	9.85	70	42	17.30	90	52	25.96	110	58	35.96	120	74
RRAU	4.90	45	32	9.90	58	37	15.5	70	42	22.5	87	60

Figs. 2–4, display the comparison of RRAU approach for JCT, MS, and MC for 15, 25, 35 and 45 resources with the conventional methodologies. The proposed algorithm is evaluated with NSGA-II [06], MOPSO [06], SPEA2 [06] and PAES [06] methods. In terms of JCT, MS and MC, RRAU approach is competed by MOPSO Algorithm. MOPSO Algorithm [6] is defined for minimizing continuous functions, where the implementation is bearable, computationally cheap and compressed. The algorithm initially performs mutation on the entire population and then it rapidly decreases its coverage over time. The technique is helpful in terms of preventing premature convergence due to the existing local Pareto fronts in some optimization problem. But, the method failed to identify a diverse set of solutions and in converging near the true pareto-optimal set. NSGA-II [6] addressed to take care of the asset allocation issue for its capacity of finding diverse set of solutions. The technique characterized another chromosome structure, pre-arranged initial population dependent on the job size and the processing speed of the assets to adjust the estimations of all objectives in resulting ages. However, the uncertainty of the asset costs, vitality utilization of assets is not considered while allotting assets. PAES [6] algorithm portrayed about gathering the stopping rules, fulfilling information dependency and reducing the complete vitality utilization, Make Span and the communication cost for shifting number of tasks and assets. But, in PAES, a solitary parent creates a single offspring in combination with a historical file that records the non-dominated solutions. This procedure expands the computational complexity. SPEA2 [6] clarified an idea for finding or approximating the Pareto- optimal set for multi-target optimization issues which moved towards a non-dominated arrangement solution and it was propelled by a natural evolution and the population evolutionary algorithms. But, SPEA2 is more computationally costly to execute the assets with large scale applications.

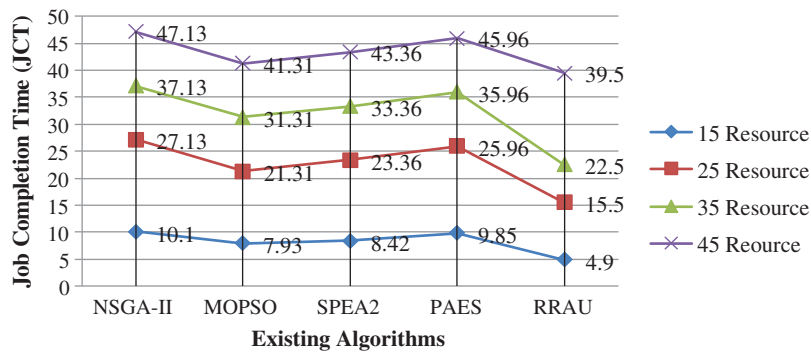


Figure 2: JCT for 15, 25, 35 and 45 resources

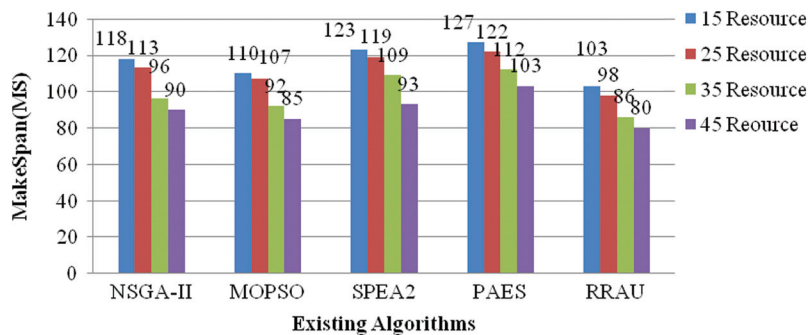


Figure 3: MS for 15, 25, 35 and 45 resource

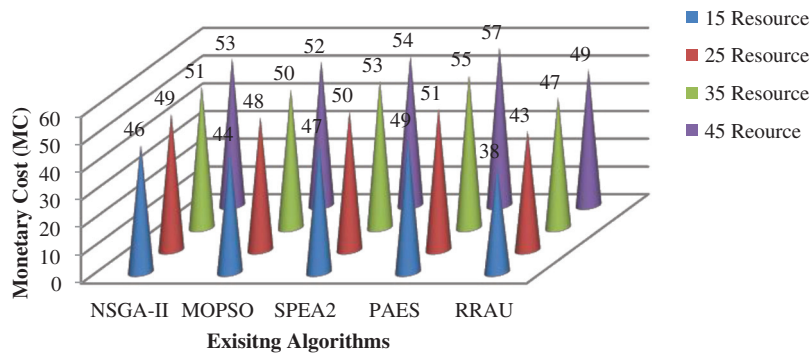


Figure 4: MC for 15, 25, 35 and 45 resource

RRAU approach optimizes the management of its cloud resources. The work hosts numerous virtual assets that would be prudent and it implements a controlled degree of QoS. The system claims that the approach is more optimal resource efficient and the servers can execute a given workload in less time. RRAU approach illustrates that the methodology is an optimal asset which is more effective than the servers and it can execute for a given outstanding task in less time. The system predicts the runtime of VMs dependent on metadata accessible at start-up. It shows that the usage of labels, which are freely-typed pieces of content used to portray VMs and altogether it improves characterization prediction results. The method evaluates and dissects the sensitivity of

a VM placement algorithm and it forecasts of VM runtimes to optimize asset and time in cloud environment. RRAU approach and J48 classifier reduces JCT 4.75 s, MS 6.25, and MC 4.25 for 15, 25, 35 and 45 resources in cloud environment. Based on the tabular and the graphical result, it can be said that the proposed algorithm performs better than the existing methods.

5 Conclusion

The article presents RRAU approach to predict the runtime of VMs dependent on metadata accessible at start-up. The method evaluates and dissects the sensitivity of a VM placement algorithm and forecasts VM runtimes to optimize asset and time in cloud environment. RRAU approach expresses to estimate under the conjecture of perfect classifications. Cloud datacenters offer a huge pool of assets available on-request. RRAU approach foresees changes in asset usage and it takes expert pro-active decisions. The method evaluates dissects the sensitivity of a VM placement algorithm and forecasts of VM runtimes to optimize asset and time in cloud environment. RRAU Approach and J48 classifier reduces JET 4.75 s, MS 6.25, and MC 4.25 for 15, 25, 35 and 45 resources in cloud environment.

In future, the work can be extended to optimize resource utilization and task scheduling in fog computing environment, where source and destination node are not reliable and job transmission creates congestion. Hence, optimal resource utilization with machine learning model can be considered to optimize the resources and schedule task frequently in fog computing.

Acknowledgement: The authors would like to thank the Doctoral committee members of Anna University, Chennai, for their valuable input and feedback. Meantime, the authors like to extend their hearty thanks to the in charge of Research Centre, RMK Engineering College too for yielding the resources.

Funding Statement: The author(s) have not received specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] R. Moreno-Vozmediano, S. Rubén, E. H. Montero and M. L. Ignacio, "Efficient resource provisioning for elastic cloud services based on machine learning techniques," *Journal of Cloud Computing: Advances, Systems and Applications*, vol. 8, no. 5, pp. 1–18, 2019.
- [2] F. Haouari, A. EmnaBaccour, M. Amr and G. Mohsen, "QoE-aware resource allocation for crowd sourced live streaming: A machine learning approach," in *2019 IEEE Int. Conf. on Communications*, Shanghai, China, pp. 1–6, 2019.
- [3] A. K. Bashir, A. Rajakumar, B. Shakila, R. Gunasekaran, J. Ramkumar *et al.*, "An optimal multi-tier resource allocation of cloud RAN in 5G using machine learning," *Transactions on Emerging Telecommunications Technologies*, vol. 30, no. 8, pp. 1–22, 2019.
- [4] Z. Hu, D. Li and G. Deke, "Balance resource allocation for spark jobs based on prediction of the optimal resource," *Tsinghua Science and Technology*, vol. 25, no. 4, pp. 487–497, 2020.
- [5] T. L. D. Thang, G. L. Rafael, C. Paolo and P. O. Östberg, "Machine learning methods for reliable resource provisioning in edge-cloud computing: A survey," *ACM Computing Surveys (CSUR)*, vol. 52, no. 5, pp. 1–39, 2019.
- [6] M. Afrin, A. R. JiongJin, T. Yu-Chu and K. Ambarish, "Multi-objective resource allocation for edge cloud based robotic workflow in smart factory," *Future Generation Computer Systems*, vol. 97, pp. 119–130, 2019.

- [7] F. Zafari, B. Prithwish, K. L. Kin, L. Jian, S. Ananthram *et al.*, “Resource sharing in the edge: A distributed bargaining-theoretic approach,” arXiv preprint arXiv: 2001, pp. 1–12, 2020.
- [8] G. Mandal, D. Santanu, D. G. Kousik and D. Paramartha, “A linear regression-based resource utilization prediction policy for live migration in cloud computing,” *Algorithms in Machine Learning Paradigms*, vol. 870, pp. 109–128, 2020.
- [9] M. Wajahat, “Cost efficient dynamic management of cloud resources through supervised learning,” *ACM SIGMETRICS Performance Evaluation Review*, vol. 47, no. 3, pp. 28–30, 2020.
- [10] I. A. Elgendy, Z. Weizhe, T. Yu-Chu and L. Keqin, “Resource allocation and computation offloading with data security for mobile edge computing,” *Future Generation Computer Systems*, vol. 100, pp. 531–541, 2019.
- [11] P. Yu, F. Zhou, X. Zhang, X. Qiu and M. Cheriet, “Deep learning-based resource allocation for 5G broadband TV service,” *IEEE Transactions on Broadcasting*, vol. PP, no. 99, pp. 1–14, 2020.
- [12] V. Podolskiy, M. Michael, K. Abigail, G. Michael and P. Panos, “Maintaining SLOs of cloud-native applications via self-adaptive resource sharing,” in *2019 IEEE 13th Int. Conf. on Self-Adaptive and Self-Organizing Systems*, Umea, Sweden, pp. 72–81, 2019.
- [13] B. Sniezynski, N. Piotr, W. Michal, J. Marcin and Z. K. Zielinski, “VM reservation plan adaptation using machine learning in cloud computing,” *Journal of Grid Computing*, vol. 17, no. 4, pp. 797–812, 2019.
- [14] Z. Chen, H. Jia and M. Geyong, “Learning-based resource allocation in cloud data center using advantage actor-critic,” in *2019 IEEE Int. Conf. on Communications*, Shanghai, China, pp. 1–6, 2019.
- [15] G. Gao, X. Mingjun, W. Jie, H. He, W. Shengqi *et al.*, “Auction-based VM allocation for deadline-sensitive tasks in distributed edge cloud,” *IEEE Transactions on Services Computing (Early Access)*, IEEE, p. 1, 2019.
- [16] K. Aziz, Z. Dounia and B. Mostafa, “Leveraging resource management for efficient performance of Apache Spark,” *Journal of Big Data*, vol. 6, no. 1, pp. 1–23, 2019.
- [17] A. Tchernykh, S. Uwe, T. El-ghazali and B. Mikhail, “Towards understanding uncertainty in cloud computing with risks of confidentiality, integrity, and availability,” *Journal of Computational Science*, vol. 36, pp. 1–9, 2019.
- [18] Q. Zeng, D. Yuqing, K. L. Kin and H. Kaibin, “Energy-efficient radio resource allocation for federated edge learning,” arXiv preprint arXiv: 1907.06040, pp. 1–14, 2019.
- [19] Y. Wei, K. Daniel, L. Shijun, P. Li, W. Lei *et al.*, “A reinforcement learning based auto-scaling approach for SAAS providers in dynamic cloud environment,” *Mathematical Problems in Engineering*, vol. 2019, pp. 1–11, 2019.
- [20] A. R. Arunarani, D. Manjula and S. Vijayan, “Task scheduling techniques in cloud computing: A literature survey,” *Future Generation Computer Systems*, vol. 91, pp. 407–415, 2019.