iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** XI **Month of publication:** November 2021

DOI: https://doi.org/10.22214/ijraset.2021.39058

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

Minimizing the Cost for Resource Allocation from Multiple Cloud Providers

Manasa Jonnagadla[1] A. Sandhya Rani[2]
[1]M.tech Student, [2]Head of the Department, Department of Computer Science and Engineering, Ananthalakshmi Institute of Technology and Sciences, A.P

Abstract: Cloud computing provides streamlined tools for exceptional business efficiency. Cloud service providers typically offer two types of plans: reserved and on-demand. Restricted policies provide low-cost long-term contracting, while order contracts are expensive and ready for short periods. Cloud resources must be delivered wisely to meet current customer demands. Many current works rely on low-cost resource-reserved strategies, which may be under- or over-provisioning. Resource allocation has become a difficult issue due to unfairness causing high availability costs and cloud demand variability. That article suggests a hybrid approach to allocating cloud services to complex customer orders. The strategy was built in two stages: accommodation stages and a flexible structure. By treating each step as an optimization problem, we can reduce the overall implementation cost while maintaining service quality. Due to the uncertain nature of cloud requests, we set up a stochastic Optimization-based approach. Our technique is used to assign individual cloud resources and the results show its effectiveness.
Keywords: Cloud computing, Resource allocation, Demand

I. INTRODUCTION

These days, many businesses prefer to host applications on cloud computing platforms. Startups use public cloud infrastructures to deploy their apps, lowering their initial costs. Larger companies are also adopting clouds, either public clouds to expand existing infrastructures or quickly deploy test environments, or private clouds for internal divisions to share virtual resources. The cloud system's resources are supposed to be pay-per-use. Workload for each task is likely multidimensional. As a result of multidimensional execution, the compute resources required may be multi-attribute (CPU, disc reading speed, network bandwidth, etc.). Second, a task may be split into multiple sequential execution phases, each requiring a different computing ability and price on demand, resulting in a potentially high-dimensional execution scenario. Online proposals pay for regulated cloud Pay-as-yougo utility distributors and encourage users to start rather than finish examples without penalty. With-order resources are much more expensive than those allocated with local intention; web services provide cases in advance with future economic deals. Sustainable investment will be difficult in quiet situations due to the complexities of quiet plans. The funds purchased may not have met the criteria, resulting in a shortage. In an uncertain economic climate, cloud and network service providers may struggle to achieve Quality of Service () [9]. However, too much provisioning can result in unnecessary costs when financially stable supplies are often unrealistic for managing actual emerging needs[1]. Cloud resource management has recently been defined, but cloud web service providers are now required to deliver data [9]-[11]. Many current cloud resource allocation algorithms design resource allocations in one step [12-15]. Writers ignore market ambiguity while recommending probabilistic market values. So cloud-based systems' dynamic existence is unknown. Demand has multiple dynamic resource distribution algorithms built-in[16]-[19]. These methods are flexible and widely spread cloud services to reduce energy distribution costs. These operations do not maximise the cost savings from cloud vendors' reserved contracts. They will then be unable to respond financially. The studies propose a hybrid approach to assigning specialised computing resources to create cloud-based mobile apps. Using allocated and ordered resources, as well as a scalable strategy to reduce production costs. The article's main findings are summarised as follows: Proposing a probabilistic system of improvement as free variables for design customer orders, and achieving a 10% overall price increase. A proposed DCRA mechanism was tested in Amazon Web Services [7] and Microsoft Azure [6] using cloud distributors and two benchmark workloads. The results show that the proposed DCRA approach recognises alternatives under uncertain customer requirements and minimises implementation costs.

II. RELATED WORK

The issue of cloud asset provisioning has carried specialists' consideration regarding giving asset portion calculations and methods in the previous few years. Most asset allotment works to frame the issue like single-stage calculation that takes into account of assets with reserved cloud suppliers' plans. To adapt to this problem, on-request asset provisioning strategies are proposed to allow assets as indicated by the dynamic cloud request.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

Deterministic Resource Provisioning Jiao et al. presented a cost model for the implementation of the online social network[13]. The analysis was planned to optimize the monitoring price of accessing cloud resources when running QoS, such as latency in access. The event that the area of the client is changed, the first arrangement will no longer meet the QoS specification. Expenditure was optimized using such a heuristic algorithm; however the answer could only be a local optimum. Similarly, a multiobjective framework[21] is proposed to reduce total deployment costs and to maximize QoS performance at the same time, while being opposing goals. This approach requires a single stage optimization that only takes into account the cloud resources assigned to the application. Imai et al. recommended agnostic performance modelling for a large application class in [12]. To improve forecast accuracy, the model is given a probabilistic predictive capability to predict application performance. To meet a Service Level Agreement (SLA), such as a latency violation, the maximum user throughput is considered. Every one of these studies suggests deterministic methods to model resource allocation issues that can lead to under- and over provision. This article proposes a dynamic cloud asset assignment algorithm for complex customer requirements.

A. Dynamic Resource Provisioning

An algorithm to reduce costs for a wide range of applications has been proposed. An Optimum Cloud Resource Provisioning (OCRP) algorithm is proposed to reduce overall service delivery costs. This algorithm has three modules[13]. If the obtained resources do not meet the specifications, additional resources will be purchased on demand. However, because the ambiguity model is discrete, the OCRP algorithm's accuracy suffers. Using the Robust Cloud Resource Provisioning (RCRP) algorithm, Chaisiri et al. Three-step cloud provider RCRP proposed. The study uses three sources of volatility: demand, costs, and availability. During the first phase, the applicant is given a sum of money. On has seen the use of available capital to consider which facilities are under- or oversupplied. If the allocated resources do not meet the targets, additional resources must be delivered in on-demand facilities. The results show that the RCRP algorithm delivers low on-demand costs compared to current works. The proposed RCRP process requires a fatalistic approach, resulting in high costs. Moreover, such a technique does not distinguish between computational and network storage issues. Throughout [22], Ran et al. proposed a low-cost, high-availability Internet productivity solution. This study's goal appears to be to reduce capital costs and maximise gains while accounting for demand and cost complexity. The proposed method uses stochastic programming in two stages. Second, determinism is expressed to reduce the cost of acquiring Virtual Machines (V.M.). The capital storage facility maximised profit in the second stage. To reduce lodging costs, Yu et al. proposed a synthetic cloud distribution approach [23]. To ensure that renters' expectations for bandwidth were met, a synthetic deterministic ensemble will be proposed The goal is to reduce resource usage in [20]. This work also investigated probabilistic linear optimization for cloud resource requirements. Unlike OCRP and RCRP algorithms that use pessimistic methods, We propose stochastic optimization to find a balanced arrangement and avoid unnecessary costs due to negative arrangements. Finally, unlike the work proposed in [20], our proposed DCRA algorithm addresses complex cloud asset allocation from a provider's perspective

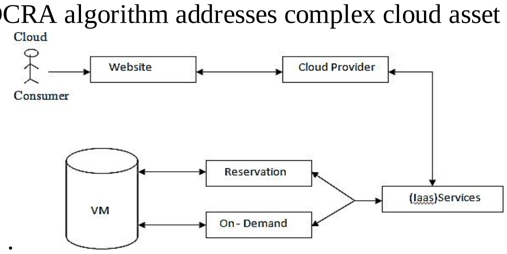


Figure1: Work Flow

III. SYSTEM DESIGN

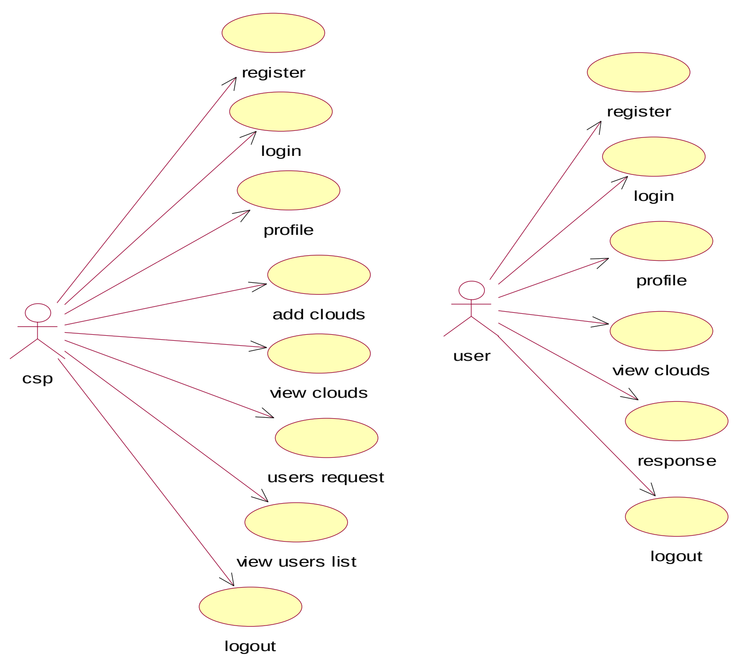
A. UML Diagrams

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group. The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML. The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems. The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems. The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

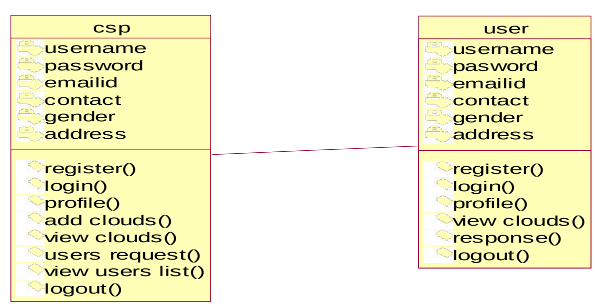
B. Goals

The Primary goals in the design of the UML are as follows:
Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
Provide extendibility and specialization mechanisms to extend the core concepts.
Be independent of particular programming languages and development process.
Provide a formal basis for understanding the modeling language.
Encourage the growth of OO tools market.
Support higher level development concepts such as collaborations, frameworks, patterns and components.
Integrate best practices.

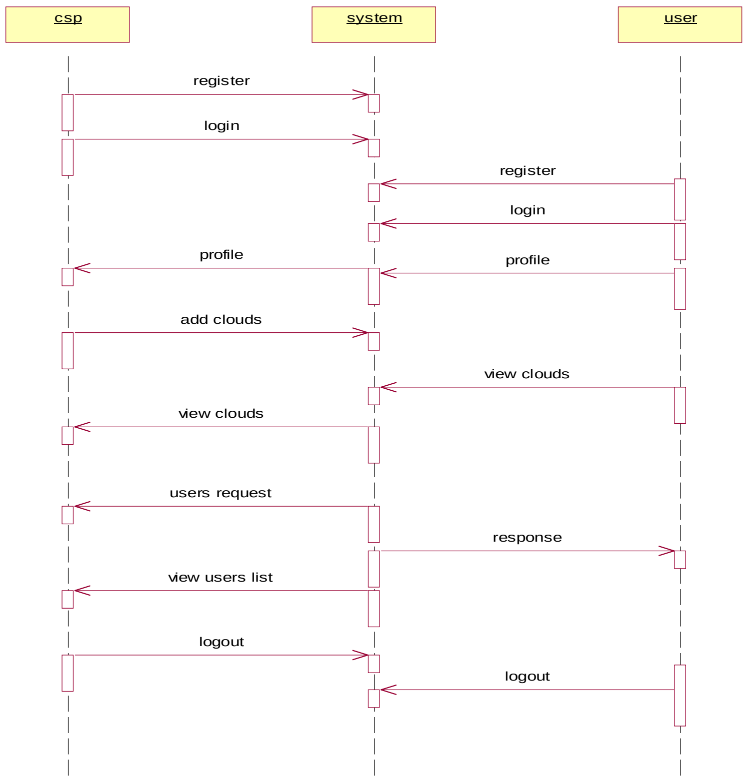
1) *Use Case Diagram:* A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



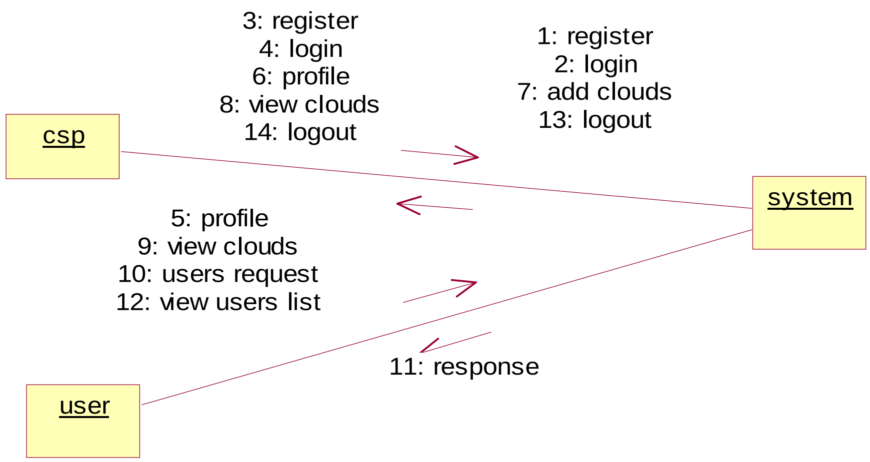
2) *Class Diagram:* In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.



3) *Sequence Diagram:* A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

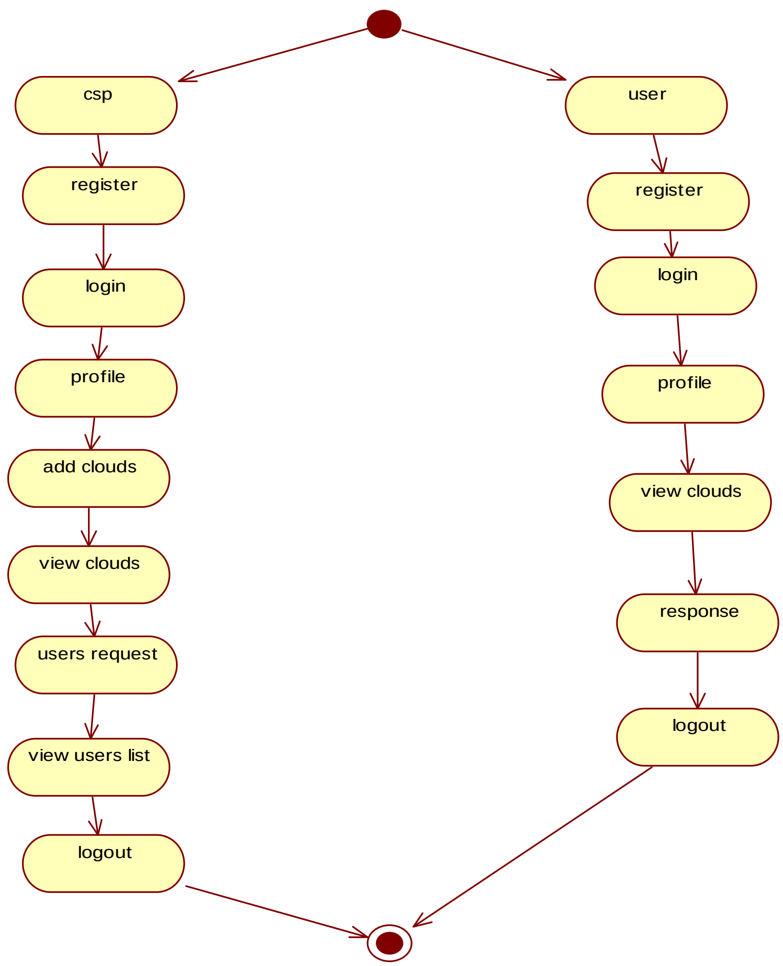


4) *Collaboration Diagram:* In collaboration diagram the method call sequence is indicated by some numbering technique as shown below. The number indicates how the methods are called one after another. We have taken the same order management system to describe the collaboration diagram. The method calls are similar to that of a sequence diagram. But the difference is that the sequence diagram does not describe the object organization whereas the collaboration diagram shows the object organization.

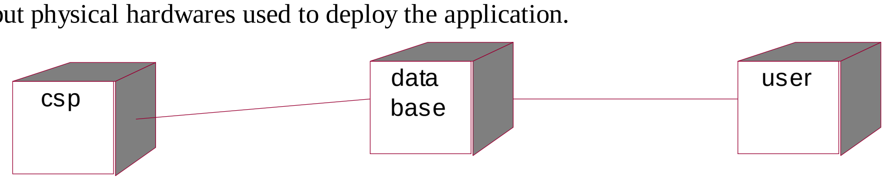


International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

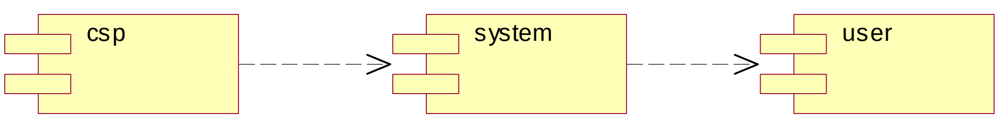
5) *Activity Diagram:* Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.



6) *Deployment Diagram:* Deployment diagram represents the deployment view of a system. It is related to the component diagram. Because the components are deployed using the deployment diagrams. A deployment diagram consists of nodes. Nodes are nothing but physical hardwares used to deploy the application.



7) *Component Diagram:* Component diagrams are used to describe the physical artifacts of a system. This artifact includes files, executable, libraries etc. So the purpose of this diagram is different, Component diagrams are used during the implementation phase of an application. But it is prepared well in advance to visualize the implementation details. Initially the system is designed using different UML diagrams and then when the artifacts are ready component diagrams are used to get an idea of the implementation.



C. Dynamic Allocation of Cloud Resources

Mostly the Cloud Vendors rely on long duratedContracts(on a yearly stand).On-demand services, on theother hand, are obtained for some arbitrary time frame, normally in hours. Reserved services are known to have a single pay whereas on-demand resources are paid during thetime period they are used. The price of these resources depends based on the length of the subscription, the long term or the short term. As shown, the Amazon pricing model[7] and the GoGrid[24] pricing model are seen inTable 1. For instance, in nearly all cloud vendors, Amazon and Azure, assets retained are held safe by long-distance contracts. In the other hand, on-demand assets are procured for every self-assertive timeframe, and on a routine basis, assets are paid for the timeframe they are used.

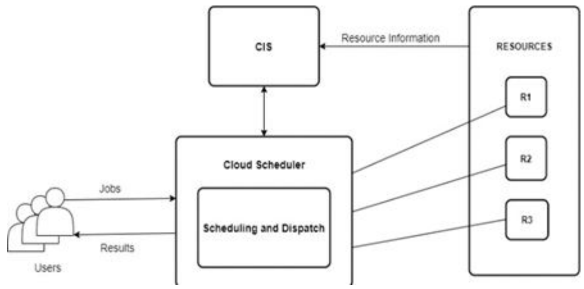


Figure 2: Web service deployment Cloud

IV. DYNAMIC CLOUD RESOURCEALLOCATION ALGORITHM

This section proposes a two-phase calculation that reduces web server usage. To meet the minimum QoS requirements, the Reservation Plan administrations are saved for web application execution in the first stage (Reservation). Non-deterministic client determinations are displayed as arbitrary factors in the second stage (complex arrangement stage).

A. DCRA Flowchart Overview

Fig. 2 depicts the reservation and complex provision phases of allocation. Cloud vendors sell online service providers certain services in the form of long-term contracts at lower costs. Like Amazon EC2 [7] reserved packages, Go Grid [24] offers yearly and monthly contracts.

Algorithm 1 Monte Carlo Simulations

Input: Cloud resource optimization solution, r_{ave} and σ
Output: #demandViols and #QoSViols

```
set #experiments 0
set #demandViols 0
set #QoSViols 0
while #experiments <= 1000 do
    generate a random demand following a normal
     distribution with mean r_ave and std. dev. σ
    if web service provider demand constraints violated
then
        #demandViols++
    if web service provider QoS constraints violated
then
        #QoSViols++
    #experiments++
return #demandViols and #QoSViols
```

The algorithm sets the required capital to fulfil the minimum customer request rmin. The web service provider only provides rmin once at the start of the implementation service period, and it is never measured or updated. These conclusions are based on employee-based data on resource allocation and usage. Even with fluctuating demand, proposed DCRA will reduce web app deployment costs.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

B. Uses of DCRA (Dynamic Cloud Resource Allocation) Algorithm

Cloud computing is a new distributed commercial computing model that aims at providing computational resources or services to users over a network in a low-cost manner.

Resource allocation and scheduling (RAS) is the key focus of cloud computing, and its policy and algorithm have a direct effect on cloud performance and cost.

This presents five major topics in cloud computing, namely locality-aware task scheduling; reliability-aware scheduling; energy-aware RAS; Software as a Service (SaaS) layer RAS; and workflow scheduling.

These five topics are then classified into three parts: performance-based RAS; cost-based RAS; and performance- and cost-based RAS.

A number of existing RAS policies and algorithms are discussed in detail accordingly with regard to their given parameters. In addition, a comparative analysis of five identified problems with their representative algorithms is performed. Finally, some future research directions of cloud RAS are pointed out.

Our assessment can use any cloud infrastructure. However, we use AWS and Azure to obtain cloud tools for web application deployment. The proposed DCRA algorithm is written in C++. The proposed technique uses the Mosek 6.0[27] solver to overcome the optimization issues. The simulation setup is being built in C++. The arrival rate of customer demand is modelled in [25]. Cloud provider output data were used to model cloud resource behaviors. The DCRA or QCost cloud asset optimization approach is used as input parameters. Because client needs for a web application are arbitrary factors, the applications cruel and standard deviation are regularly considered.

1) Advantages

a) The main advantage of resource allocation is that users do not need to install software or hardware to access, develop, or host applications online.
b) There is no limitation of place or medium. We can access our applications and data from any system.
c) The user saves money on hardware and software.
d) Insufficient resources, cloud providers can share resources online.

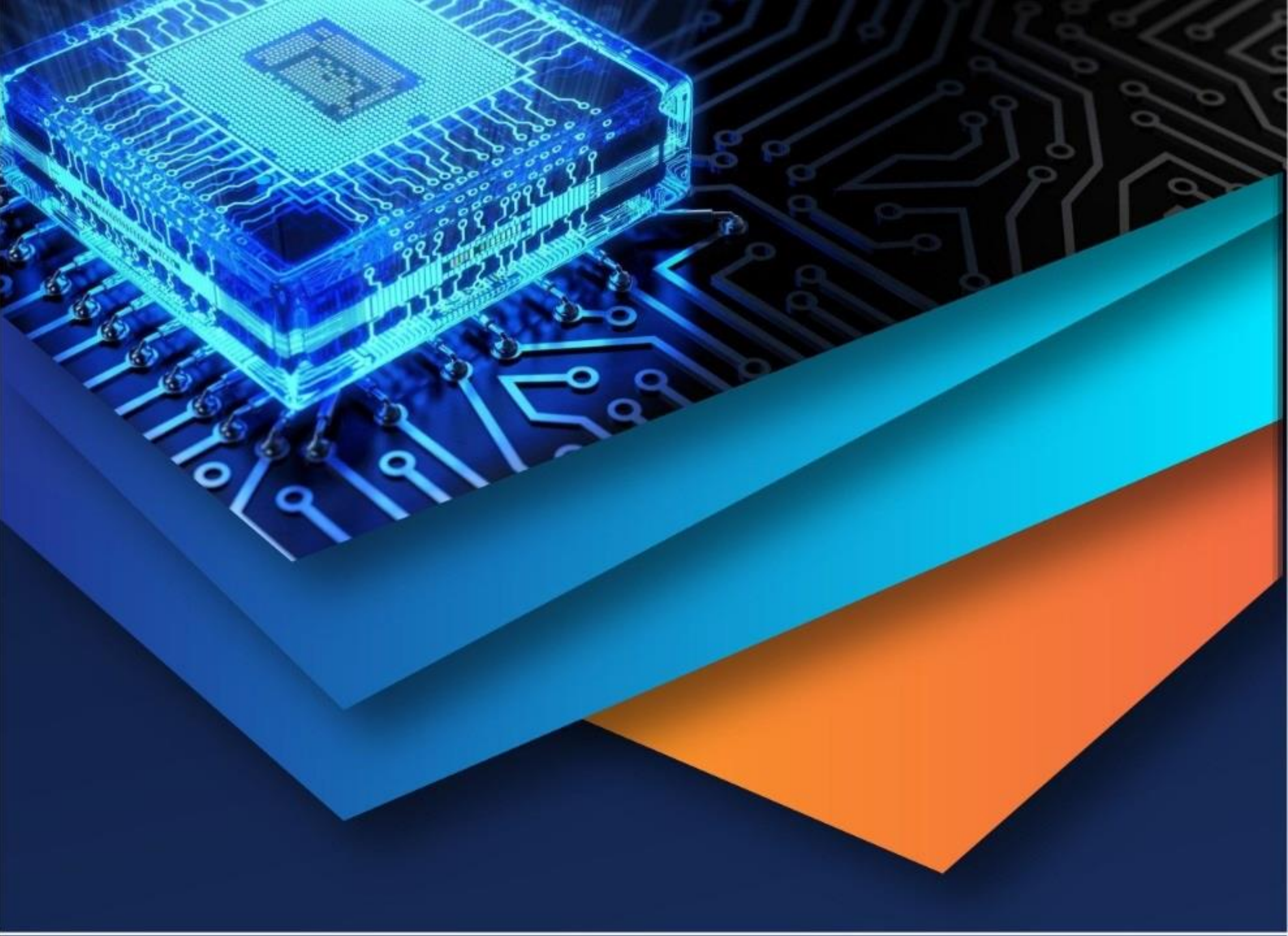
V. CONCLUSION

As a result, a hybrid provisioning algorithm is proposed that focuses on maximising the overall amount of provisioning that can be obtained, while taking into account the shopper requirements' volatility. First, there will be a Reservation phase; then, there will be a Dynamic Provisioning phase. Re-enactments of various workload scenarios were used to gauge DCRA's performance. For cloud-based applications, it appears that the proposed method can provide cost-effective and effective solutions by combining distributed and on-demand services. When compared to AWS [7], Azure [6], and Go Grid [24], the proposed DCRA is a self-contained cloud service. Expanding on DCRA's existing capabilities, the proposed project will automate the distribution of services among the various cloud-based applications. It is possible that priority scheduling techniques will be employed in the future when allocating client resources.

REFERENCES

[1] Mireslami, S., Rakai, L., Wang, M., & Far, B. H. (2019).” Dynamic Cloud Resource Allocation Considering Demand Uncertainty”. IEEE Transactions on Cloud Computing, 1–1.
[2] M.Anastasopoulos, A.Tzanakaki, and D. Simeonidou, "Stochastic energy-efficient cloud service Provisioning deploying renewable energy sources," IEEE Journal on Selected Areas in Communications,vol. 34, no. 12, pp. 3927–3940, Dec 2016.
[3] A. Johannes, N. Borhan, C. Liu, R. Ranjan, and J. Chen, "A user demand uncertainty based approach for cloud resource management," in 2013 IEEE 16th International Conference on Computational Science and Engineering, Dec 2013, pp. 566–571.
[4] S. Hosseinalipour and H. Dai, "Options-based sequential auctions for dynamic cloud resource allocation," in 2017 IEEE International Conference on Communications (ICC), May 2017, pp. 1–6.
[5] Rackspace. https://www.rackspace.com.
[6] Microsoft Azure. http://azure.microsoft.com/
[7] Amazon web services. http://aws.amazon.com/.
[8] AWS Amazon EC2. https://aws.amazon.com/ec2/pricing/on-demand/.
[9] S.Chaisiri, B.S.Lee, and D. Niyat, "Robust cloud resource provisioning for cloud computing environments," in 2010 IEEE International Conference on Service-Oriented Computing and Applications (SOCA), Dec 2010, pp. 1–8.
[10] N.Sfika, A.Korfiati, C.Alexakos, S. Likothanassis, K.Daloukas, & P. Tsompanopoulou, "Dynamic cloud resources allocation on multidomain/multiphysics problems," in 2015 3rd International Conference on Future IoT and Cloud, Aug 2015, pp. 31–37.

[11] R.I.Meneguette,A.Boukerche, A.H.M.Pimenta, & M.Meneguette, "A resource allocation scheme based on semi-Markov decision process for dynamic vehicular clouds," in 2017 IEEE International Conference on Communications (ICC), May 2017, pp.1–6.
[12] S.Imai, S.Patterson, & C.A.Varela, "Cost-efficient elastic stream processing using application-agnostic performance prediction," in 2016 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), May 2016, pp. 604–607.
[13] L. Jiao, J. Li, T. Xu, W. Du, and X. Fu, "Optimizing cost for online social networks on geo-distributed clouds," IEEE/ACM Transactions on Networking, vol. 24, no. 1, pp. 99–112, Feb 2016.
[14] H. Goudarzi, M. Ghasemazar, and M. Pedram, "SLAbased optimization of power and migration cost in cloud computing," in the 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), May 2012, pp. 172–179.
[15] W. K. Tan, D. M. Divakaran, and M. Gurusamy, "Uniform price auction for allocation of dynamic cloud bandwidth," in 2014 IEEE International Conference on Communications (ICC), June 2014, pp. 2944–2949.
[16] J. Chase and D. Niyato, "Joint optimization of resource provisioningin cloud computing," IEEE Transactions on Services Computing, vol. PP, no. 99, pp. 1–1, 2015.
[17] S. Chaisiri, B. S. Lee, and D. Niyato, "Optimization of resource provisioning cost in cloud computing," IEEE Transactions on Services Computing, vol. 5, pp. 164–177, April 2012.
[18] J. N. Khasnabish, M. F. Mithani, and S. Rao, "Tiercentric resource allocation in multi-tier cloud systems," IEEE Transactions on Cloud Computing, vol. 5, no. 3, pp. 576– 589, July 2017.
[19] B.Neethu and K.R.R.Babu, "Dynamic resource allocation in market oriented cloud using auction method," in 2016 International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE), Sept 2016, pp. 145–150.
[20] M.Anastasopoulos, A. Tzanakaki, and D. Simeonidou, “Stochastic energy efficient cloud service provisioning deploying renewable energy sources,” IEEE Journal on Selected Areas in Communications,vol. 34, no. 12, pp. 3927–3940, Dec 2016.
[21] S. Mireslami, L. Rakai, B. H. Far, and M. Wang, “Simultaneous cost and qos optimization for cloud resource allocation,” IEEE Transactions on Network and Service Management, vol. 14, no. 3, pp.676–689, Sept 2017.
[22] Y. Ran, B. Yang, W. Cai, H. Xi, and J. Yang, “Cost-efficient provisioning strategy for multiple services in distributed clouds,”in 2016 International Conference on Cloud Computing Research and Innovations (ICCCRI), May 2016, pp. 1–8.
[23] L. Yu and H. Shen, “Bandwidth guarantee under demand uncertainty in multi-tenant clouds,” in 2014 IEEE 34th International Conference on Distributed Computing Systems, June 2014, pp. 258–267.
[24] GoGrid. https://www.datapipe.com/gogrid/.
[25] X. Nan, Y. He, and L. Guan, “Optimal Resource Allocation for Multimedia Cloud Based on Queuing Model,” in the 13th IEEE International Workshop on Multimedia Signal Processing (MMSP), Hangzhou, China, October 17-19 2011, pp. 1–6.
[26] A. Dastjerdi, S. Garg, and R. Buyya, “QoS-aware deployment of network of virtual appliances across multiple clouds,” in IEEE International Conference on Cloud Computing Technology and Science(CloudCom), November 2011, pp. 415–423.
[27] Mosek 6.0. http://www.mosek.com.



10.22214/IJRASET

INDEX COPERNICUS

45.98

ISRA JIF

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY