



Emerging computational challenges in cloud computing and RTEAH algorithm based solution

Srinivasa Rao Gundu¹ · Charan Arur Panem² · Anuradha Thimmapuram³ · R. S. Gad⁴

Received: 10 June 2020 / Accepted: 5 July 2021 / Published online: 14 July 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

The new era of computation has given new directions for advancement and sophistication. The meaning of computation has made a paradigm shift from device and location orientation to the distributed level. Cloud has working models with a service-oriented delivery mechanism as well as a deployment oriented infrastructure mechanism. Data centers are the backbone of cloud computing. Multi-cloud exchanges allow organizations to establish a single connection to multiple cloud providers at the same time through an Ethernet switching platform, rather than wrestling with multiple individual connections to cloud providers. The massive participation of the public has also increased the load on the cloud servers. Proper scheduling of resources is always needed. There are many load balancing algorithms are there and all of them have their limitations. Different circumstances need different load balancing approaches. One single algorithm would not be able to provide a proper scheduling mechanism in all the needed directions. Therefore ‘hybrid algorithms’ would be best suited to provide improved services to the needs of the present conditions. This present paper analyzed a few and this can be carry forwarded to the next level with some other algorithms and parameters with vigorous efforts to analyze.

Keywords Azure Data Factory · Azure DevOps · Blob · Cloud computing · Microsoft Azure · Pipeline · Virtualization · Web services

1 Introduction Need to change here

Earlier in the 1940s, IBM started Mainframes. Mainframes are large computational units leveraging multiple processing. These systems are used for processing huge amounts of data such as ERP (Enterprise Resource Planning) applications. Mainframes run continuously without any interruption.

Because of emerging technologies modifications in main-frame applications were not easy (Abbas et al. 2020).

The computational model in the 1960s has taken a new dimension by introducing ‘Job control commands’. Job control commands are used in batch processing mode. It has minimized the human interactions between any consecutive computational tasks by avoiding the ‘idling state’.

In 1969, ‘Leonard Weinrock’ started the internet as part of ARPANET (Advanced Research Projects Agency Network). Internet is a network of networks. The Internet has opened doors for many more technologies to bloom. The below Fig. 1. Shows how different systems are connected across the United States of America (Ahmed and Ahmed et al. 2018).

Cloud computing technology has succeeded from the grid concept. Cloud computing systems are generally established in bigger data centers. These data centers would be hosted by a company. Cloud is interactive technology (Tripathi et al. 2017).

Software as a service (SaaS) satisfies the software requirements, Platform as a service (PaaS) for applicational needs, and Infrastructure as a service (IaaS) for

✉ Srinivasa Rao Gundu
srinivasarao.gundu@gmail.com

Charan Arur Panem
panem.charan@gmail.com

Anuradha Thimmapuram
rajamata@yahoo.co.in

R. S. Gad
rsgad@unigoa.ac.in

¹ Department of Computer Science, Dravidian University, Andhra Pradesh, India

² Department of Electronics, Goa University, Goa, India

³ Dravidian University, Andhra Pradesh, India

⁴ Goa University, Goa, India

Fig. 1 Advanced Research Projects Agency Network Project MAP which connects the computer systems across the United States of America in 1969

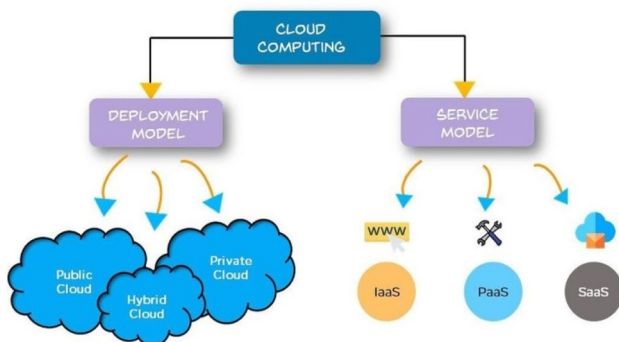
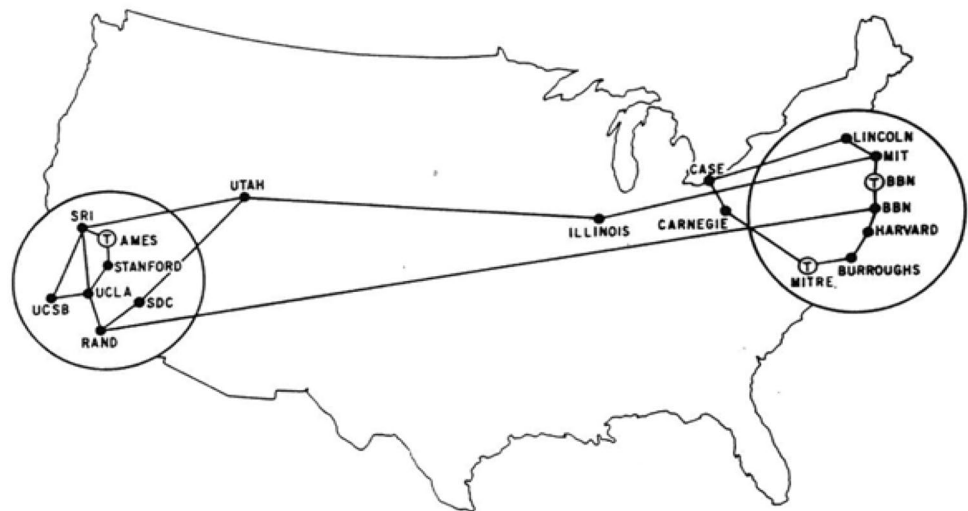


Fig. 2 Different models of cloud computing

infrastructures. These infrastructures include virtual machines. The above-mentioned services can be had with a variety of deployment categories (Fig. 2).

Cloud computational model is intended to provide efficient utilization of computational resources and to reach the end-user. Now it is possible to define the cloud with a societal terminology such as “Cloud is a technology for the user, by the service providers and to the user” (Drakopoulos et al. 2019). Cloud’s technological advancement has increased the user’s computational capacity by not merely restricted to a specific computational system.

2 Aim of the study

This present study aims to re-evaluate the emerging computational trends and techniques and provide an optimal solution to the existing problems.

3 Organization of the paper

The organization of the paper is as follows:

Title of the Paper
Abstract
Keywords

1. Introduction
2. Aim of the study
3. Organization of the paper
4. Literature Review
5. Cloud computing and Data Centers
6. Resources allocation in Cloud
7. Problem Statement
8. Mathematical Model
9. Few selected load balancing algorithms
10. Hybrid load balancing algorithm in cloud
11. Research Methodology: Proposal of a hybrid algorithm approach
12. RTEA Hybrid algorithm: Analysis
 - (a). Pseudocode
 - (b). Flow Chart
 - (c). Simulation tool for the Investigation- Cloud sim
 - (d). Simulation Environment
 - (i). Space-shared policy
 - (ii). Time-shared policy
 - (e). Experimental setup
 - (f). Stepwise Description of CloudSim Data Flow
 - (g). Inputs supplied
 - (h). RTEAH Algorithm Observations

Conclusion
 Future Scope
 Acknowledgements
 Statement of Funding
 Conflict of Interest Statement
 Acknowledgment
 Funding Statement
 Conflict of Interest
 References

4 Literature review

As per Kim (2019), there is excessive growth of various sorts of Cyber-Physical Systems. There are only a few studies that recognized the complete Cyber-Physical Systems. His study has performed a complete evaluation of the recent Cyber-Physical Systems research relevant to upcoming Information technology (IT) developments. He has addressed above mentioned four distinct technologies in terms of integration, verbal exchange, network management, control, and safety.

As per Yang (2019), genomics Next Generation Sequencing (NGS) and Third Generation Sequencing (TGS) have a vast vicinity of applications in life science, such as Non-Invasive Prenatal Testing (NIPT), ctDNA Testing for Non-Invasive Tumor Personalized Therapy. In his study, he has given the design and implementation of a genomics cloud, which can scale storage and computing abilities flexibly and provides many easy to use genomics analysis software for customers. As per Aceto et al. (2020), Industry 4.0 and its main enabling information and communication technologies are completely changing both services and production worlds. In his study, he has provided (i) a description of the main technologies and paradigms concerning Healthcare 4.0 and discuss (ii) their main application scenarios; we then provide an analysis of (iii) carried benefits and (iv) novel cross-disciplinary challenges; (v) the lessons learned.

As per Mitra et al. (2017), one time password (OTP) scheme has been suggested as an efficient and simple solution for authentication in cloud computing. Application of linear rules in synthesis of One time password (OTP)'s ensures easy implementation using only modular arithmetic, which makes the proposed method truly cost efficient. Cost efficiency and design flexibility in generation of one time password (OTP) sets have been achieved with the proposed Equal Length Cellular Automata (ELCA)-based approach.

As per He (2014), the advancements in cloud computing and internet of things (IoT) have provided a promising opportunity to resolve the challenges arised by the increasing transportation issues. He presented a novel multilayered vehicular data cloud platform by using cloud computing and IoT technologies.

As per Xie (2017), Information resource services are the important element for resource sharing in cloud manufacturing. In this study, a semantic model is presented for data resource service modeling that uses semantic links instead of ontologies. This model takes advantage of semantic links to enable automated integrating and distributed updating in the resource service cloud.

As in keeping with Zheng et al. (2014), cloud computing is an critical aspect of the backbone of the Internet of Things (IoT). Clouds could be required to aid large numbers of interactions with various quality necessities. In his study CLOUDQUAL incorporates six high-quality dimensions, i.e., usability, availability, reliability, responsiveness, security, and elasticity, of which usability is subjective, whereas the others are goal.

As per Lu (2020), cloud computing is an evolutionary technology that offers on-demand resources and elastic services through the Internet. Most providers adopt pay-as-you-go pricing mechanisms. This study proposes a novel auction method that can efficiently schedule the computational resources as per the customer's Quality of Service (QoS) requirements.

As per Fang (2017), integrated information systems are important safeguards for the utilization and development of land resources. In this study, an economical and highly efficient supervision system for land resources has been established based on IoT and cloud computing technologies; a novel online and offline integrated system with synchronized internal.

As per Liu et al. (2014), cloud manufacturing has recently become a hot research topic in the manufacturing industry. One of the key bottlenecks that hinders the development and application of cloud manufacturing is security. To enable authorized users in the right location, right time and right network to access resources, a GeoAuthentication model that maps geographic location, access time, and sub network information into a secret key is proposed.

As per Li (2012), this article presents a hybrid wireless network integration scheme in cloud services-based enterprise information systems (EISs). As a key feature, the proposed approach integrates access control functionalities within the hybrid framework that provide users with filtered views on available cloud services based on cloud service access requirements and user security credentials.

As per Xu et al. (2018), Rapid advances in industrialization and informatisation methods have spurred tremendous progress in developing the next generation of manufacturing technology. In this paper, it is briefly surveyed the state of the art in the area of Industry 4.0 as it relates to industries.

As per Banan et al. (2020), the rapid emergence of deep learning (DL) technology has resulted in its successful use in various fields, including aquaculture. In brief, the purpose is to provide researchers and practitioners with a better

understanding of the current state of the art of DL in aquaculture, which can provide strong support for implementing smart fish farming applications.

As per Fan (2020), Estimation of absolute temperature distributions is crucial for many thermal processes in the nonlinear distributed parameter systems, such as predicting the curing temperature distribution of the chip, the temperature distribution of the catalytic rod, and so on. Satisfactory agreement between the results of the current model and the other well-established model shows that the KL-MLP-LSTM model is reliable for estimating the temperature distributions during the thermal process.

As per Shamshirband (2019), nowadays, learning-based modeling system is adopted to establish an accurate prediction model for renewable energy resources. Computational Intelligence (CI) methods have become significant tools in production and optimization of renewable energies. In all cases, hybrid networks have better performance compared with single networks, because hybrid techniques take the advantages of two or more methods for preparing an accurate prediction. It is recommended to use hybrid methods in DL techniques.

As per Ganjkhani (2019), this paper provides a novel bad data detection processor to identify false data injection attacks (FDIAs) on the power system state estimation. The results indicate that the proposed bad data detection processor is able to detect the false injected data launched into the system accurately.

5 Cloud computing and data centers

Cloud has provided some benefits such as, (i). It has reduced the computational costs, (ii). It is device-independent, (iii). It is location independent, (iv). It has a huge storage capacity for the users and providers maximum of up to 3000 GB, (iv). Computing power has increased (v). Software costs are reduced, (vi). Software installation and updating are possible without any hazards for the users (Amanatullah et al. 2013).

Cloud computing is supported by three major technological aspects. They are, (i). Virtualization shares the resources among the users. (ii). Service-Oriented Architecture (SOA) facilitates applications as a service to the end-user. Irrespective of vendors or technology applications are possible to avail.

The data center and cloud architectures are kept on advancing, to address the requirements of expansive scale multi-server farms in clouds. These requirements are revolved around seven measurements called (i) scalability, (ii) storage, (iv) bandwidth, (v) speed in-network services, (vi) efficiency in memory utilization, (vii) agility in service creation, (viii) cost productivity. The below are a few advantages of cloud computing (Ayadi et al. 2013).

Today, more enterprise organizations are moving toward a Hybrid IT infrastructure to increase their agility, reduce the total cost of ownership, and create overall positive business outcomes. IT teams need adequate time to ensure interoperability across all cloud providers along with advanced knowledge in areas like container technology, cloud network interconnections, and management of cloud-native micro services which may not be feasible for every team to achieve, alongside their daily responsibilities (Duan et al. 2015).

Looking at the practical grounds, and from the Cloud Datacenter Organizers point of view there are four different types of challenges are there. They are (i). Common Challenges will arise at the Infrastructure Level in cloud data centers, (ii). Challenges that arise with cloud service providing Competitors, (iii). Performance and providing Quality of service to the customers, (iv). The growth of load on Servers. Generally, a Data Center usually consumes approximately 0.1 MW to 15 MW. Power Usage Effectiveness (PUE) is used to estimate the power usage in Data Centers (Maenhaut et al. 2017).

Even though Moore's Law says about the increment of performance of transistors every approximately 730 days, in the case of power consumption there is no much improvement is seen. Since the Data Centers emit much more heat, there is a need to cool to provide an environment-friendly data center.

Nowadays CRAC units (Computer Room Air Conditioning Units) are used to look guard the temperature maintenance, distribution of air in the Data Center, and the humidity level to control. Cloud computing resources are a collection of computational resources and network resources.

These resources are the real assets of cloud computing. Memory, CPU time, processing time, processing speed, disk storage, bandwidth, etc. are a few to mention here. These computational resource usage can be measured partially by Big O notation. An algorithm is engineered in such a way to optimized recurring processes or an optimized repetitive process of computational and network resources. Therefore flexibility is to be achieved in the algorithm (Wang et al. 2017).

6 Resources allocation in cloud

As cloud computing is a collection of heterogeneous environments, the nature of the user-requests also reflects. The user requests need different computational aspects and they vary from time to time. Peak hours and non-peak hours also matters a lot to show the difference in the resources requirements.

The resources need to be scheduled as per the needs and accordingly. Scheduling algorithms are numerous in

literature and they have their advantages and disadvantages within their limits. Therefore it is impossible to provide a Quality of Service without proper scheduling of computational resources.

Cloud services are provided to the end-users based on its service-oriented architecture, which is provided with the help of the internet. The huge success of cloud computing is possible with the help of virtualization technology. Virtualization technology helps to slice the data center to act like many numbers of servers. Even though it depends on the hardware configuration of the data center in how many virtual machines can be divided, to implement virtualization additional software is also required. Such software is commercially available (Duraksha et al. 2019).

The Quality of Services which is the most desired and agreed point of view of both service providers and the end-user depends on several points, in which one of which is load balancing. The load refers to the amount of data traffic being carried by the cloud computing network. The load balancing mechanism is completely dependent on the amount of work allotted to the system for a specific period. Proper Scheduling of tasks would reduce the problem of starvation. There are many load balancing algorithms of different types and categories, but all of them had their limitations. Although it is not possible to make the best load balancing algorithm, improvement in this regarding is always open.

A computer system needs scheduling before use; the CPU is one of the most critical parts of the computer. Multiprogramming is one of the basic and important scheduling techniques. Generally, CPU scheduling is done in such a way as to keep it busy as much as possible.

Scheduling occurs due to any one of the four conditions such as the given below (i) When a process went from switching state to waiting for state or (ii) When a process is terminated or (iii) When a process switches from running state to the ready state or When a process switches from waiting for the state to a ready state. In cases (i) and (ii) there is no choice for scheduling and this is non-preemptive scheduling. Here a new process is to be selected for the execution. It is also called cooperative scheduling. Under this scheduling, if one Central Processing Unit is allocated to a process, the process keeps the Central Processing Unit busy until it releases the Central Processing Unit or by terminating or by switching it to the waiting state. In cases (ii) and (iii) there is a choice for scheduling. This is pre-emptive scheduling. It is related to share data when a kernel on behalf of the processor is working and busy, 'processor waiting' is required. Rarely it could be interrupted, but it is not desirable.

Scheduling performance can be evaluated using a synthetic workload (Salapura 2012) below the parameters (i). Distribution of several tasks, (ii). Distribution of real-time job services on demand, (iii). Average inter-arrival time of tasks, (iv). Period of real-time jobs.

Scheduling policies can be stated as below (i).FCFS (First Come First Serve), (ii).Round Robin (Execution of tasks in a cyclic manner), (iii). Shortest Job First (SJF) etc.

7 Problem statement

Queuing theory and scheduling have both developed techniques to deal with dynamic resource allocation and sequencing. The scheduling approaches typically solve an NP-complete problem during each scheduling period, it is assumed that the "time pressure" of the problem is low. That is, one time unit in the scheduling model represents a reasonable amount of "real-time" and therefore it has the time to use optimization techniques to improve the schedule (Wang et al. 2016).

Queuing manage issues, on the contrary, are dynamic: the intention in such problems is traditionally to investigate the choicest action to take when the method is in an exact state. Queuing thought ordinarily reviews systems with easy combinatorics, as such systems are more amenable to rigorous analysis of their stochastic methods (Henry et al. 2018).

8 Mathematical model

The study deals with the formulation of optimization programming problems to explore the decision Parameters which regulates the Virtual Machine. Computation of different stages continuous time-dependent high capacity required user request incidence rates can be done with stochastic models. These mathematical models have the limitations that they are formulated with the assumptions of handling multiple problems independently and separately. Let A and B be two parameters that influence the Virtual Machines like throughput, makespan, etc.

Let T- Transformation ratio of normal capacity required User request into high capacity required user Request.

I-Initial number of user requests the needs less activity.

K-Initial number of user requests that need high Capacity.

Then the objective function (M) predicts the maximizing cases of high capacity required request. Let A and B be two parameters that influence the Virtual Machines like throughput, makespan, etc.

$$M = I / (T - 1/A - B) + (K - T I / A - B)$$

User request handling capacity, server capacity, and geographic location play an important role. Mathematical Modeling studies are the meaning full alternatives to handle the Virtual Machines. The model construction should be stochastic models that will provide the most relevant picture of the prevailing conditions. Stability is an important idea in

queuing conception and types the predominant part of the theoretical evaluation. Informally, a method is stable if its queues stay bounded over time.

9 Few selected basic cloud load balancing algorithms

There are many numbers of load balancing algorithms. Here are a few basic algorithms used widely and they have their limitations.

- (i). Round Robin algorithm (Basic version) is the easiest algorithm used for the logic division of time and nodes. This algorithm has a simple process with the automatic mechanism of selecting the next process in the queue. In this algorithm, the overall time is partitioned into the number of segments, and each node in the system is allocated with a particular time segment or time. Therefore it works on the logic of time division. Every node will be assigned with some time along with work. The main responsibility of a cloud service provider is accepting the request from the client, and allotting time segment for each request. But this Algorithm is static and does not maintain the present status of each node at the time of allocation (Srinivasa Rao and Anuradha 2018a). Drawbacks of the Round-robin algorithm: (a) Clients have to wait in the waiting queue until and unless a suitable virtual machine is available. (b). Additional Load on Scheduler to decide the size of quantum. (c). It has a longer average waiting time, higher context switches, higher turnaround time, and low throughput. (d). It does not consider the previous load state of a node at the time of assigning the jobs. (e). Requires prior knowledge of user tasks and system resources, does not make the use of current state of the system. (f). It fails to model the heterogeneous nature of the cloud computing environment.
- (ii). Throttled algorithm (Basic version): It is a dynamic algorithm or static scheduling algorithm, It is a virtual machine-based algorithm, Throttled virtual machine load balancer keeps a list of virtual machines and their status whether they are busy or idling when the user cloudlet request comes assigns to a proper virtual machine and lets the work to be done. Drawbacks of Throttled Algorithm: (a). It works properly only if all virtual machines in the data center have the same hardware configuration. (b). It should be made some Modifications in the existing algorithm so that it could achieve flexibility in its implementation.

- (iii) Equally Spread Current Execution Algorithm (Basic version): In the spread spectrum technique load balancer makes effort to preserve equal load to all the virtual machines connected with the data center. Equally, Spread Current Execution Load Balancing Algorithm characteristics as follows: (a). It is a Dynamic Algorithm, (b). It is a virtual machine Load Balancing Algorithm, (c). It is based on the spread spectrum technique, (d). It Loads the scheduling tasks equally i.e., equally spreading, (e). Load balancer monitors the scheduled jobs. ESC load balancer puts all the tasks in the job pool and assigns them to the virtual machine. This algorithm balances the load properly up to some extent (Srinivasa Rao and Anuradha et al. 2018b). Drawbacks of New RTH Algorithm: For updating the Index table leads to an overhead and may cause a delay in responding to arrived requests. It could find a slight improvement in comparison with RTH and ESCE.
- (iv). Artificial Bee Colony Optimization (Basic version): Karaboga, proposed a foraging behavior of honey bee swarm in 2005. The artificial Bee Colony Optimization technique was inspired by the above-mentioned behavior of the bee swarm. It is motivated by the intelligent foraging behavior of honey bees and is the simulation of the minimalistic foraging model of the honey bee in the search process for solving real-time parameter, non-convex, and non-smooth optimization problem. In this algorithm, a colony would consist of three types of artificial bees including employed, onlooker, and scout bees (Subramanian et al. 2010).

Drawbacks of Round robin algorithm: (a). It is slow to obtain an accurate solution, (b). It has the possibility of losing relevant information on the behavior of the function to be optimized, (c). It requires new fitness tasks on the new algorithm parameters to improve performance, (d). It lacks the use of secondary information about the problem.

10 Hybrid load balancing algorithm in cloud

In many times a particular single load balancing algorithm will not be approachable due to its internal characteristics, such as the average, low performance of the algorithm, scalability, response time, throughput, etc. parameters. Generally, such algorithms are modified and denoted with the latest versions, discarding the basic versions. But latest versions will not be the final solution for all the needs. Hybridization of the algorithm is also an approach to make the performance of a system improve. A hybrid load-balancing algorithm is a combination of two or more load balancing algorithms in a cloud computing environment.

The internal working mechanism of a hybrid load balancing algorithm will look like as given below.

- (I). When a user sends a ‘request’ to the load balancer, it resolves the load balancer’s domain name using a domain name server (DNS).
- (II). Domain name server returns one or more internet protocol addresses (IP address) to the client machine.
- (iii). With network load balancers, load balancing creates a network interface for each available virtual machines list which is either busy or idle
- (iv). Each load balancer node in this virtual machine list uses this network interface to get a static internet protocol address (static IP Address)
- (v). It can be optionally associated with one internet protocol address with each network interface when it creates the load balancer.
- (vi). As traffic to the application changes over time, load balancing scales the load balancer and updates the DNS entry.
- (vii). Here one important aspect is that domain name server entry specifies the TTL (Time to Live) as sixty seconds. TTL makes the internet protocol address can be re-mapped quickly in response to change the traffic.
- (viii). The client machine determines which internet protocol address needs to use for sending the requests to the load balancer.
- (ix). Load balancer node requests the server.
- (x). Load balancer node selects a virtual machine that can handle the request.
- (xi). Now the load balancer node sends the request to the target using its private internet protocol address (Private IP Address)
- (xii). When a load balancer accepts incoming traffic from the client and routes the requests to its virtual machine pool. The load balancer also checks whether the virtual machine is either idle or busy and ensures that it routes the traffic only to the virtual machines which are capable to handle it.
- (xiii). If it is being used equally spread the current execution algorithm, for example, it is cross-zone load balancing. In this load balancing, the nodes of the load balancer would distribute the requests regardless of the availability of virtual machines and distributes the traffic evenly across all the virtual machines in the pool of virtual machines.
- (xiv). With the application of a load balancer, the load balancer node receives the request and evaluates the priority order to determine which rule to apply, and then selects a virtual machine from the pool of virtual machines for the rule action using the routing algorithm.
- (xv). Routing is performed independently for each one of the virtual machines in the virtual machine pool, even though when the virtual machine is assigned to the pool of virtual machines.
- (xvi). With network load balancers the load balancer node receives the connection to select the virtual machine from the virtual machine pool for the default rule using a flow hash algorithm. Based on the protocol, source internet protocol address (Source IP Address), source port, destination internet protocol address (Destination IP Address), destination port, and transmission control protocol sequence number, select the virtual machine.
- (xvii). The transmission control protocol (TCP) connections from a client have different source ports and sequence numbers and can be routed to a different virtual machine. Each transmission control protocol connection is routed to a single virtual machine (Lopes Ferreira 2015).

11 Research methodology: proposal of a hybrid algorithm approach

The research methodology is empirical. Primary data is taken from some of the cloud providers, Rackspace, iWeb, space, and qtcloud services are a few to mention here. Various Data centers are visited in Hyderabad. Cloud4C & CtrlS-CTRLS, NettLinx, Ricoh Data Center, and National Informatics Centre (NIC)—Hyderabad has provided relevant information.

It is the main aim to provide a comparative study of the performance of efficient load balancers in the cloud data center, and from there selecting an optimal load balancer. To do this below steps are needed to be done.

Step-1: It must be looked into different techniques by comparing various load balancer’s performance. It can be looked into some parameters such as processing cost, response time, etc., to evaluate those load balancers (Piao et al. 2016).

Step-2: It must select a load balancer algorithm or propose which can use minimum resources and within a short processing time. A hybrid approach to Load balancing can be made either in two ways: (a) Software Approach, (b) Hardware Approach. In Software Approach it shall be interested in the algorithmic approach. It shall look into the scheduling algorithm for load balancing.

Load balancing is categorized into four basic types,

(i) Centralized Load balancing algorithm, (ii) Distributed Load balancing algorithm, (iii) Hierarchical Load balancing algorithm, (iv) Workflow Dependent Load balancing algorithm (Hossan et al. 2019).

12 RTEA hybrid algorithm

The working mechanism of the RTEA Hybrid algorithm will have the below characteristics.

- (I). The Data Center controller receives the new user cloudlet request randomly from the round-robin algorithm.
- (II). RTEAH Load balancer keeps a list of virtual machines and their status whether they are busy or idling from the index table
- (III). All tasks in the job pool assign the virtual machine as the load balancer monitors the scheduled jobs, balancer keeps track of jobs that are in the queue frequently for new tasks assigns the pool of virtual machine pool of jobs assigns to a pool of virtual machines
- (IV). Calculate the load of the user cloudlet request and calculate the capacity of virtual machines.

A new hybrid approach based load balancing algorithm can be had as RTEAH Algorithm. Here H is Hybrid.

- (a). Pseudocode
Begin

Step: 1: Hybrid Load Balancer maintains an Index of virtual machines and their status whether The virtual machines are busy or idle. In the beginning, none of the Virtual Machines are allocated to any of the user requests.

Step: 2: The Data Center controller receives the user requests i.e., U_r (Basically they are Cloudlets) Let U_r enter the Data Center controller receives the user requests/Cloudlets. Calculate the load of the user request. Calculate the load and the capacity of a Virtual Machine. Let all Virtual Machines are checked in the circular way. The user requests are allocated to Virtual Machines based on their statuses which are known from the Virtual Machine Queue: if the status is available then cloudlet is allocated. The Hybrid Load Balancer will allocate the time Quantum for user request execution

Step: 3: After the execution of cloudlets the Virtual Machines are de-allocated by the Hybrid Load balancer.

Do.
Find the next available virtual machine
Update and Load the Capacity of Virtual Machine
Calculate Virtual Machine Future resource and value of each Machine. Choose the low loaded

Machine and migrate the task from the overloaded machine

If $U_r1 \leq VM3$

Assign the U_r1 to $VM3$

Assign the Request Queue/Pool to the VM Queue

While

Step: 4: The Data Center controller checks for new/pending/waiting requests in Queue

Count the active load on Virtual Machine

Update the Load and the Capacity of Virtual Machine Submit the ID of Those Virtual machines which are consisting of Minimum Load If $U_r > VM1$

Swap the U_r1 to $VM2$ If $U_r1 > VM2$

Swap the U_r1 to $VM3$.

Step: 5: Continue from Step-2

Termination Criterion is not violated

- (b). Flow Chart (Fig. 3).

- (c). Simulation tool for the Investigation- Cloud sim

When a particular algorithm is designed and developed, it needs to be tested in a simulation environment. Cloud sim is such an environment that gives the

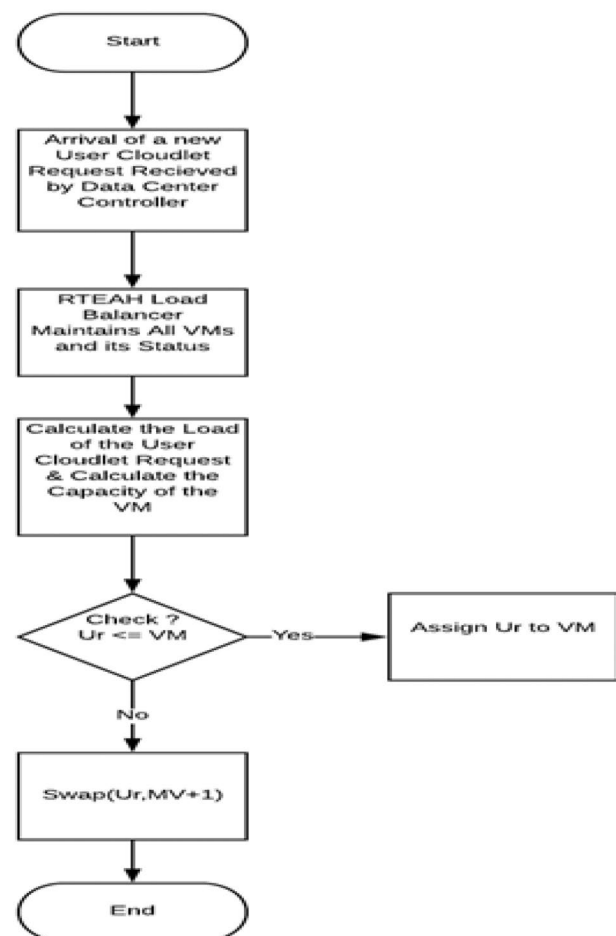


Fig. 3 RTEAH Algorithm working mechanism

user to test his/her web-oriented applications to test. To make such a test, it needs to change any number of parameters and check again and again. Cloud sim can be used to model data centers, hosts, service brokers, scheduling, and allocation policies of a large-scaled cloud platform.

Every Datacenter should register with the Cloud Information Service registry (CIS). The user requests are mapped to suitable cloud providers. The available list of cloud providers is issued by the CIS (Chang et al. 2012).

(d). Simulation environment

There are two scheduling strategies defined, They are Time-shared scheduling policy and Space-shared scheduling policy. These scheduling strategies can be applied to both VMs and task units. CloudSim allows two-level VM provisioning: firstly, at the host level, and secondly, at the VM level.

(i). Space-shared policy:

Step 1: Accepted task units are arranged in the ready queue.

Step 2: Find out if there are any free processing cores available or not.

Step 3: If available, then task units are put in the execution queue.

Step 4: First task unit is assigned to the hosting VM.

Step 5: After completion of the first task, the next task is assigned.

Step 6: If the queue is empty, it checks for the new task.

Step 7: Then it repeats from step 1.

Step 8: End.

Note 1: As each VM requires two cores, only one VM can run at a given instance of time. Only after completion of VM1, VM2 can be assigned the cores.

Note 2: Each task unit requires only one core.

(b). Time-shared policy:

Step 1: Accepted task units are arranged in the ready queue.

Step 2: All the task units are simultaneously assigned to VM. These task units are dynamically context switched during their life cycle.

Step 3: When the queue is empty, it checks for the new task.

Step 4: Then it repeats from step 1.

Step 5: End.

Note: Each VM gets a time slice completion of VM1, VM2 can be assigned the cores (Space-shared). Therefore, during a VM lifespan, all the assigned task units are dynamically context switched during their life cycle (Time-shared).

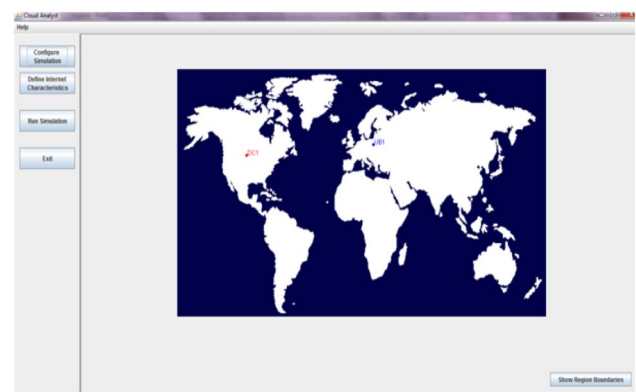


Fig. 4 Cloud Analyst tool Initial Page

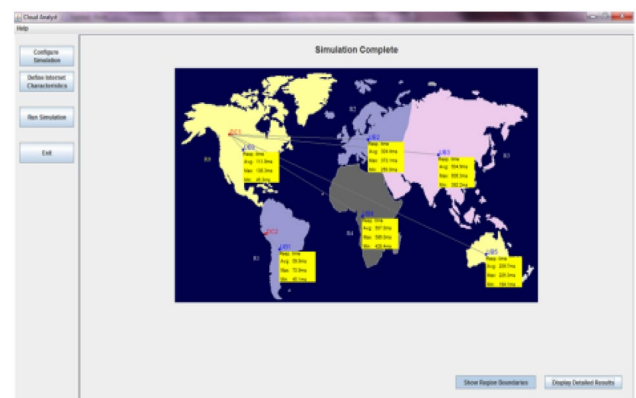


Fig. 5 Cloud Analyst tool Simulation running state -2

This processing happens according to a policy (time-shared or space shared), defined by the cloudlet scheduler. The host can submit cloudlets to the VM to be executed. According to the requirement, the following inputs have been taken to create the VMs and cloudlets (Dong-xing 2012).

(e). Experimental setup

Virtual Machine Load balancing algorithms can be implemented using the following soft wares and tools. Operating system: Windows XP or Windows 7., IDE: Net beans 8.1/Eclipse-Kepler, Toolkit: JDK 1.8, and Cloud Analyst tool. Algorithms are implemented for the IaaS model in a simulated cloud environment. Cloud Analyst is an extension to CloudSim oriented simulator used for modeling and simulation of the real cloud environment (Figs. 4, 5, 6, 7, 8, 9, 10, 11).

(f). Step wise description of cloudsim data flow

Step 1: Cloud user sends the job request to the User Interface.

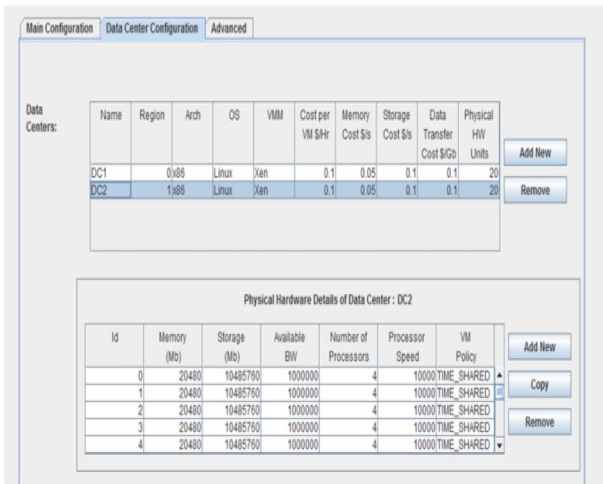


Fig. 6 Cloud Analyst tool simulation configuring the data center

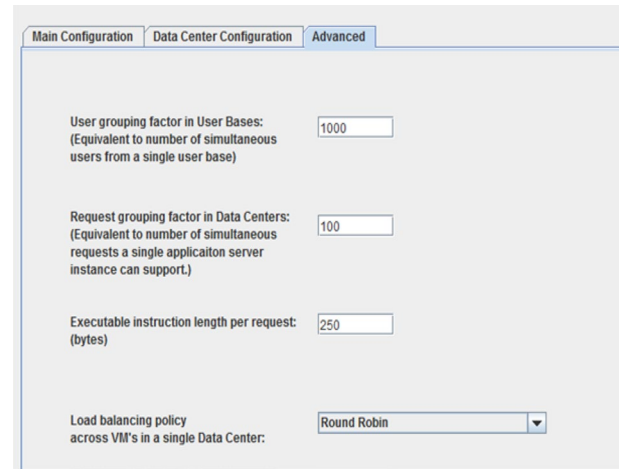


Fig. 8 Cloud Analyst tool simulation selecting the load balancing policy

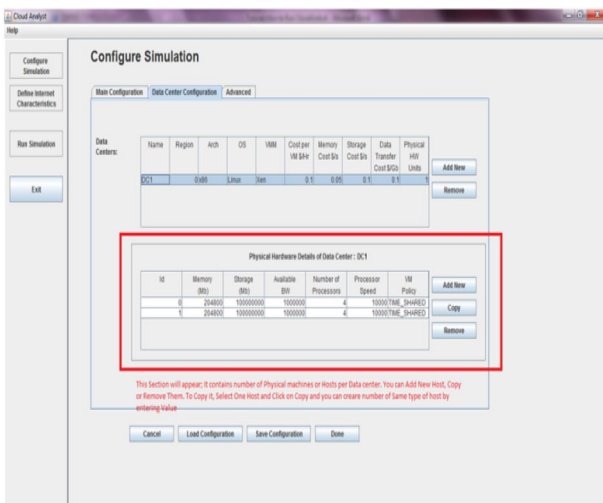


Fig. 7 Cloud Analyst tool simulation providing the physical hardware details of the data center



Fig. 9 Cloud Analyst Simulation Completed

Step 2: User Interface analyzes the request according to the Service Level Agreement (SLA).

Step 3: User Interface assigns the tasks to the cloud broker. Cloud broker is also termed as Data Center broker.

Step 4: Cloud broker (Data Center broker) divides tasks into same sized cloudlets. Cloudlet is also termed as task units.

Step 5: Cloud broker sends the cloudlets to Virtual Machine Manager (VMM).

Step 6: Each data center entity registers with the Cloud Information Service (CIS) registry.

Step 7: Cloud broker sends the resource request to Cloud Information Service (CIS). Then Cloud broker consults the CIS to get hold of the record of resources

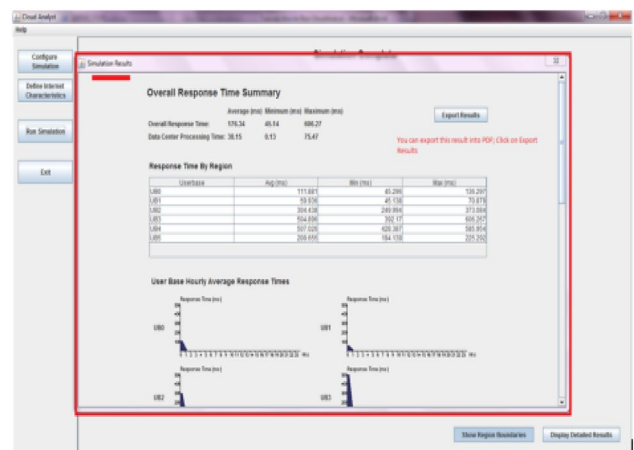


Fig. 10 Cloud Analyst tool simulation Over all Response time Summary

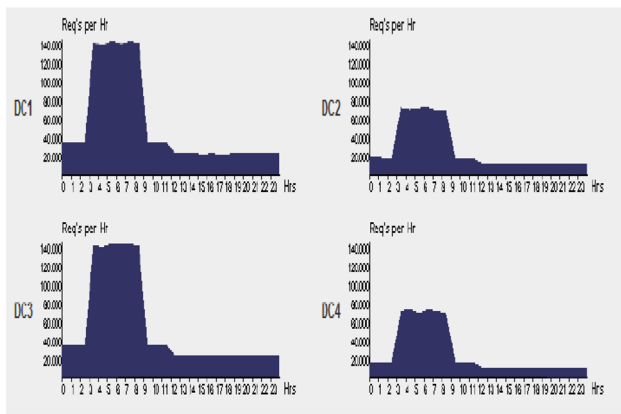


Fig.11 Data Center loading times in Cloud Analyst tool simulation

which can recommend hardware services that match the user’s infrastructure requirements.

Step 8: The cloud broker gets information about the availability of the data center and resources from the CIS.

Step 9: Virtual machine manager (VMM) creates the virtual machine.

Step 10: Data Center entity invokes UpdateVM-Processing for every host that manages it because concerned VMs manage each task unit. That’s why the processes should be updated.

Step 11: At the viewpoint of the host, invocation of UpdateVMProcessing triggers a UpdateCloudlet-Processing process that orders every Virtual Machine (VM) to update its task unit status (executing, finish, suspend) with the data center entity.

Step 12: VM analyzes the approximate execution time and sends it to the host machine.

Step 13: The host machine analyzes the smallest time to the next event.

Step 14: Datacenter provides the information about the execution time and different resources such as Operating System, VMM used, RAM size, MIPS, a number of cloudlets, the number of CPU, storage capacity, etc.

Step 15: Request for the execution of the cloudlet is sent to the virtual machine by VMM.

Step 16: Cloudlet is being executed in the VM.

Step 17: VM sends the executed cloudlets to the VMM.

Step 18: After completing the execution, VM releases the resources for further use.

Step 19: CIS updates the registry according to the information sent by the data center.

Step 20: VMM further passes the executed cloudlets to the cloud broker.

Step 21: Cloud broker combines all the executed cloudlets together to form the task.

Step 22: Cloud broker sends the completed task to the User Interface.

Step 23: After completion of the task, User Interface can either expire the session or make another renewal request.

Step 24: If the session is expired, then User Interface sends the executed task to the cloud user.

(g). Inputs supplied:

Virtual machine Configuration

Virtual machine memory—1024 Mb, RAM – 204,800 Mb,

Storage space – 100 TB, Bandwidth—10,0000, VM scheduling policy—Time shared policy 4 1000 MIPS.

Data Centre Configuration

Data Centre Architecture – X86, Data Centre Operating System – Linux, Data Centre virtual machine manager – Xen, Number of machines in the data center – 20, Memory per machine in the data center – 2048 Mb, Storage per machine in data center –100,000 Mb, Executable instruction length in the data center– 250 (Tables 1, 2, 3, 4, 5, 6, 7, 8).

Researcher’s algorithms (improvement in percentage).

(h). RTEAH algorithm observations

1. It is observed as following.
2. Case: 1. If a Simple Web Application Hosted on a Single Data Center Response time could be expected to rise much higher.
3. Case: 2. When a Web Application Hosted on Multiple Data Centers around the World.

Table 1 Available bandwidth between the regions measured in Mbps

SL. No.	Region name	R ₀	R ₁	R ₂	R ₃	R ₄	R ₅
1	R ₀	20	10	10	10	10	10
2	R ₀	10	80	10	10	10	10
3	R ₀	10	10	25	10	10	10
4	R ₀	10	10	10	10	10	10
5	R ₀	10	10	10	10	50	10
6	R ₀	10	10	10	10	10	20

Table 2 Delay matrix and transmission between the regions in nano sec

SL. No.	Region name	R ₀	R ₁	R ₂	R ₃	R ₄	R ₅
1	R ₀	20	10	10	10	10	10
2	R ₀	10	80	10	10	10	10
3	R ₀	10	10	25	10	10	10
4	R ₀	10	10	10	10	10	10
5	R ₀	10	10	10	10	50	10
6	R ₀	10	10	10	10	10	20

Table 3 Data center configuration details

SL. No	Region name	R ₀	R ₁	R ₂	R ₃	R ₄	R ₅
D.C -1	r ₀	20	10	10	10	10	10
D.C -2	R ₀	10	80	10	10	10	10
D.C -3	R ₀	10	10	25	10	10	10
D.C -4	R ₀	10	10	10	10	10	10
D.C -5	R ₀	10	10	10	10	50	10
D.C -6	R ₀	10	10	10	10	10	20

Table 4 Physical hardware details

ID No	Availbl. Mem	Storg. Capacit	Availble B/W	Number of process	Proces. Speed	Virtual machine policy
1	204,800 MB	100,000,000 GB	1,000,000 Mbps	4	10,000 Mbps	Time Shared Policy

Table 5 User base properties

Name of the Scenario	No. of User bases	Regions	Request User thr	Req.size	PeakHr start GMT	PeakHr. End at GMT	Avg. No.of Peak Users	Avg. No.of OFF Peak Users
Scenario -1	1	0	60	20	3	9	1000	100
Scenario -2	3	2,4,5	60	20,30	3	9	1000	100
Scenario -3	6	1,2,4	60	1,100,20,20	3	9	1000	100
Scenario -4	4	1,3,4	60	1,100,20,30	3	9	1000	100
Scenario -5	5	1,2,5	60	1,100,20,20	3	9	1000	100
Scenario -6	3	0.5	60	1,20,40,50	3	9	1000	100

Table 6 RTEAH Algorithm simulation results

Resp. Time	Exe. Time	Turn arnd Time	Proc. Time	Thrhg put	NW delay	Make span	Wtng. time	Burst time
292.47	0.03004	0.24	0.23	3.309	11.44	292.46	0.01	0.23

4. Case: 3. If there are two Data Centers with 25 VMs in each, then the average response time reduced (This is an Overloaded VMs case).
5. Case: 4. If there are two Data Centers with 50 VMs in each, then the average response time is improved.
6. Case: 5. If there are two data centers with 50 Virtual Machines each, and in case they are sharing the load during the peak hours, in terms of region-wise change is not observed but the data center loading patterns changes are observed.
7. Case: 6. On applying the throttled algorithm to the processing of requests Data center loading patterns changed.
8. Case: 7. A web application hosted on 3 Datacenters with 50 Virtual Machines each Overall response time is improved.

Table 7 RTEAH Algorithm in comparison with other Researcher's algorithms

Name of the parameter	Name of the Researcher	Imprvmt of RTEAH Algo. in nano sec	RTEAH Imprv.mnt (%)
Response time	Agraj Sharma	259.03	23
Execution time	Kiryong Ha	0.2	27
Turnaround time	Ritu	0.76	38
Processing time	Mrs.Snehal, A.Narale	0.44	24
Throughput	Peng Zhang	0.3	4
Network delay	Andrew, Addison	38.56	39
Make span	Jion feng, Zhao	19.54	3
Waiting time	Tanuja, Zha	0.09	45
Burst time	Tanuja, Zha	2.77	46

Table 8 RTEAH Algorithm in comparison with other Researcher's algorithms (improvement in percentage)

Name of the parameter	Name of the researcher	Other researcher's value	RTEAH algorithm (Proposed)
Response time	Agraj Sharma	551.5	292.47
Execution time	Kiryong Ha	0.5	0.03004
Turnaround time	Ritu	1.0	0.24
Processing time	Mrs.Snehal A.Narale	1.14	0.23
Throughput	Peng Zhang	3.6	3.309
Network delay	Andrew Addison	50.0	11.44
Make span	Jion feng Zhao	312.0	292.46
Waiting time	Tanuja Zha	0.1	0.01
Burst time	Tanuja Zha	3.0	0.23

9. Case: 8. A web application hosted on 3 Data Centers with 75, 50, and 25 Virtual Machines in each, overall response time considerably reduces.

(i). Results from the observations

Service quality can be increased by the application of load balancing at the application level and also at virtual machine level with data center. Any improvement in load balancing algorithms employment must be effective to meet the peak demand.

13 Conclusions

For smaller data centers and less traffic load static load balancing algorithms are best suited but among them, some other parameters also need to consider such as more power consumption. This is one step towards establishing green clouds. Meanwhile, if the algorithm takes more computational resources then it must be addressed with some improvement or by considering another algorithm. Hybrid IT mixes co-located on-premise and cloud-based IT to combine the security and performance of on-premise facilities with the public cloud's agility, scale, and cost savings. As part of a Hybrid IT strategy, organizations can spread workloads and apps such as CRM, ERP, and others across multiple clouds for greater flexibility, redundancy,

cost management, and security. RTEAH Load balancer keeps a list of virtual machines and their status whether they are busy or idling in the form of an index table, all tasks in the job pool assigns the virtual machine as the load balancer monitors the scheduled jobs, balancer keeps track of jobs which are in the queue frequently for new tasks assigns the pool of virtual machines pool of jobs assigns to a pool of virtual machines, calculate the load of the user cloudlet request and calculate the capacity of virtual machines. The New RTEAH algorithm has shown improvement in case our cloud computing load balancing parameters in terms of response time, Execution time, Turnaround time, Processing time, Throughput, Network delay. Therefore RTEAH is highly considerable.

14 Future scope

There are many number of algorithms are there for the improvement of load balancing in cloud computing, but all of them has their own drawbacks (Table 1). Proposed approaches can be extended to the other parameters also such as Disk space usage, Central Processing Unit (CPU) Processing Speed, Latency, Fault tolerance.

Acknowledgements I wish to thank various Data centers in Hyderabad Cloud4C & CtrlS-CTRLS, NettLinx, Ricoh Data Center, and National Informatics Centre (NIC)—Hyderabad who has guided me to collect relevant information. I sincerely thank my Research Supervisor and Ex-Vice-Chancellor (i/c) of Dravidian University, Prof. T. Anuradha for her elderly contribution to this paper and for her guiding through her teaching, research guidance which made me explore more into the qualitative content about the cloud computing environment. I wish to thank Late Mr. Panem Nadipi Chennaih for his continues support for the development of this research paper and it is dedicated to him.

Funding This Study is not funded by any organization.

Declarations

Conflict of interest We do not have any conflict of Interest.

References

- Abbas I, Ahmad M, Faizan M, Arshed W and Khalid J (2020) Issues and challenges of cloud computing in performance augmentation for pervasive computing. In: 2020 international conference on electrical, communication, and computer engineering (ICECCE), Istanbul, Turkey, pp 1–7
- Aceto G, Persico V, Pescapé A (2020) Industry 4.0 and health: Internet of things, big data, and cloud computing for healthcare 4.0. *J Indust Inform Integrat* 18(2):100–129
- Ahmed EA, Ali Ahmed H (2018) A proposed model for education system using cloud computing. In: 2018 3rd international conference on emerging trends in engineering, sciences and technology (ICEEST), Karachi, Pakistan, pp 1–4

- Ali D, Gupta MK (2019) Deadline sensitive lease scheduling using hungarian genetic algorithm in cloud computing environment. *Internat J Comput Sci Eng* 7(12):7–15
- Amanatullah Y, Lim C, Ipung HP, Juliandri A (2013) Toward cloud computing reference architecture: Cloud service management perspective. In: 2013 International Conference on ICT for Smart Society, Jakarta, pp 1–4
- Ayadi I, Simoni N, Aubonnet T (2013) SLA approach for cloud as a service. In: 2013 IEEE Sixth International Conference on Cloud Computing, Santa Clara, CA, pp 966–967
- Banan A, Nasiri A, Taheri-Garavand A (2020) Deep learning-based appearance features extraction for automated carp species identification. *Aquacult Eng* 89(2):1–10
- Chang C, Yang J, Chang K (2012) An efficient and flexible mobile payment protocol. In: 2012 sixth international conference on genetic and evolutionary computing, Kitakushu, pp 63–66
- Dongxing H (2012) Information network security situation aware technology based on artificial intelligence[J]. *J Inform Comm* 6(2012):80–81
- Drakopoulos G, Kafza E, Al Katheeri H (2019) Proof systems in blockchains: a survey. In: 2019 4th South-East Europe design automation, computer engineering, computer networks and social media conference, PiraeusGreece, pp 1–6
- Duan L, Zhan D, Hohnerlein J (2015) Optimizing cloud data center energy efficiency via dynamic prediction of CPU Idle Intervals. In: 2015 IEEE 8th international conference on cloud computing, New York, NY, pp 985–988
- Fan YJ (2020) Spatiotemporal modeling for nonlinear distributed thermal processes based on KL decomposition. *MLP and LSTM Network IEEE Access* 8(1):25111–25121
- Fang S (2017) An integrated system for land resources supervision based on the IoT and cloud computing. *Enterprise Inform Syst* 11(1):105–121
- Ganjkhani M (2019) A novel detection algorithm to identify false data injection attack on power system state estimation. *Energies* 12(2209):1–9
- He W (2014) Developing vehicular data cloud services in the IoT environment. *IEEE Trans Industr Inf* 10(2):1587–1595
- Henry R, Herzberg A, Kate A (2018) Blockchain access privacy: challenges and directions. *IEEE Secur Priv* 16(4):38–45
- Hossain MT, Chowdhury MZ, Shahjalal M, Jang YM (2019) Human bond communication with head-mounted displays: scope, challenges, solutions, and applications. *IEEE Commun Mag* 57(2):26–32
- Kim JH (2017) A review of cyber-physical system research relevant to the emerging IT trends: industry 4.0, IoT, big data, and cloud computing. *J Ind Integr Manag* 2(3):1750011
- Li S (2012) Integration of hybrid wireless networks in cloud services oriented enterprise information systems. *Enterp Inform Syst* 6(2):165–187
- Liu X, He W, Xu L, Yan G (2014) Enhancing the security of cloud manufacturing by restricting resource access. *J Homel Secur Emerg Manage* 11(4):533–554
- Lopes Ferreira GA (2015) Internet of things and the credit card market: how companies can deal with the exponential increase of transactions with connected devices and can also be efficient to prevent frauds. In: 2015 12th international conference on information technology-new generations, Las Vegas, NV, pp 107–111
- Lu Y (2020) Pricing the cloud: a QoS-based auction approach. *Enterp Inform Syst* 14(3):334–351
- Maenhaut P, Moens H, Volckaert B, Ongena V, De Turck F (2017) Resource allocation in the cloud: from simulation to experimental validation. In: 2017 IEEE 10th international conference on cloud computing (CLOUD), Honolulu, CA, pp 701–704
- Mitra A, Kundu A, Chattopadhyay M, Chattopadhyay S (2017) A cost-efficient one time password-based authentication in cloud

- environment using equal length cellular automata. *J Ind Inf Integr* 5(1):17–25
- Piao C, Zuo Y, Zhang C (2016) Research on hybrid-cloud-based user privacy protection of O2O platform. In: 2016 IEEE 13th international conference on e-business engineering (ICEBE), Macau, pp 214–219
- Salapura V (2012) Cloud computing: virtualization and resiliency for data center computing. In: 2012 IEEE 30th international conference on computer design (ICCD), Montreal, pp 1–2
- Shamshirband S (2019) A survey of deep learning techniques: application in wind and solar energy resources. *IEEE Access* 7(1):164650–164666
- Srinivasa Rao G, Anuradha T (2018a) Improved hybrid approach for load balancing in virtual machine. *Internat J Comput Sci Eng* 6(10):730–733
- Srinivasa Rao G, Anuradha T (2018b) Improved implementation of hybrid approach in cloud environment. *Internat J Comput Sci Eng* 6(10):254–260
- Subramanian V, Wang L, Lee E, Chen P (2010) Rapid processing of synthetic seismograms using windows azure cloud. In: 2010 IEEE second international conference on cloud computing technology and science, Indianapolis, pp 193–200
- Tripathi R, Vignesh S, Tamarapalli V, Medhi D (2017) Cost efficient design of fault tolerant geo-distributed data centers. *IEEE Trans Netw Serv Manag* 14(2):289–301
- Verma A, Malla D, Choudhary AK, Arora V (2019) A detailed study of azure platform and its cognitive services. 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon). Faridabad, pp 29–134
- Wang Z, Zeng J, Lv T, Shi B, Li B (2016) Cloud auditor: a cloud auditing framework based on nested virtualization. In: 2016 IEEE 3rd international conference on cyber security and cloud computing (CSCloud), Beijing, pp 50–53
- Wang Y, Xia Y, Chen S (2017) Using integer programming for workflow scheduling in the cloud. In: 2017 IEEE 10th international conference on cloud computing (CLOUD), Honolulu, CA, pp 138–146
- Xie C (2017) Linked semantic model for information resource toward cloud manufacturing. *IEEE Trans Industr Inf* 13(6):3338–3349
- Xu L, Xu E, Li L (2018) Industry 4.0: state of the art and future trends. *Internat J Prod Res* 56(8):2941–2962
- Yang J (2019) Cloud computing for storing and analyzing petabytes of genomic data. *J Ind Inform Integr* 15(7):50–57
- Zheng X (2014) Cloud service negotiation in internet of things environment: a mixed approach. *IEEE Trans Industr Inf* 10(2):1506–1515

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.