IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 2, September 2024

Cloud Computing: Revolutionizing IT Infrastructure with On-Demand Services and Addressing Security Challenges

Suraj Patel
Automotive IT Infrastructure, Detroit, USA
surajbpatel88@gmail.com

Abstract*: Cloud computing has emerged as a revolutionary computing paradigm that integrates virtualization, parallel and distributed computing, utility computing, and service-oriented architecture. This model allows enterprises to leverage scalable and flexible IT infrastructure, reducing capital expenditures and operational costs. Cloud computing offers various services such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS), providing computing resources on-demand with a pay-as-you-go pricing model. Popular commercial implementations include Amazon's EC2, Google App Engine, and Salesforce's CRM system. While cloud computing offers immense benefits in terms of cost efficiency, scalability, and immediate time-to-market advantages, it also raises significant security concerns, particularly in data security and privacy. Ensuring data confidentiality and implementing robust access control mechanisms are crucial to addressing these security challenges. Without resolving these issues, the future widespread adoption of cloud computing could be hindered. In this paper we have result show response time for RR, ESCE, TTL and TLB for overall response time and data center processing time.*

Keywords: Cloud Computing, Virtualization, Distributed Computing, Utility Computing, Service-Oriented Architecture, Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS), Scalability, Cost Efficiency

I. INTRODUCTION

Cloud computing has gained significant traction in recent years as a cutting-edge technological paradigm that integrates various existing fields such as virtualization, distributed computing, and utility computing. It is reshaping the way enterprises and individuals consume computing resources by offering a range of services that reduce the reliance on physical infrastructure. At the core of cloud computing are its service models: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS), which provide different levels of abstraction for managing and using resources [1-2].

The key benefits of cloud computing include scalability, flexibility, and cost efficiency, as users can access resources on demand without upfront investments in hardware or software. Companies like Amazon, Google, and IBM have successfully implemented cloud services, offering flexible pricing models that allow users to only pay for what they use [3].

However, with the increasing reliance on cloud platforms, data security and privacy concerns have become major obstacles. Since data in the cloud is stored and processed on servers managed by third-party providers, users are required to entrust their sensitive information to these providers. Ensuring that data remains confidential and secure is essential for the future growth and trust in cloud computing technologies[4].

Data represents an extremely important asset for any organization, and enterprise users will face serious consequences if its confidential data is disclosed to their business competitors or the public. Thus, cloud users in the first place want to make sure that their data are kept confidential to outsiders, including the cloud provider and their potential competitors. This is the first data security requirement [5-7]. Data confidentiality is not the only security requirement. Flexible and fine-grained access control is also strongly desired in the service-oriented cloud computing model.

ISSN 2581-9429 IJARSCT

DOI: 10.48175/IJARSCT-19647

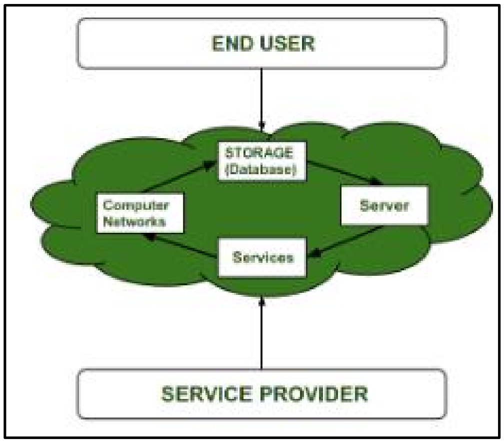


Fig. 1 Basic Roots of Cloud Computing

A health-care information system on a cloud is required to restrict access of protected medical records to eligible doctors and a customer relation management system running on a cloud may allow access of customer information to high-level executives of the company only. In these cases, access control of sensitive data is either required by legislation (e.g., HIPAA) or company regulations [7-8].

II. SERVICES OF CLOUD COMPUTING

Cloud computing is born for to provide services to consumers as per their requests, basically cloud computing services is divided into three main classes, according to abstraction level of the capability provided by service model of providers namely [9-11].

- Infrastructure as a Service i.e. IaaS,
- Platform as a Service i.e. PaaS,
- Software as a Service i.e. SaaS.

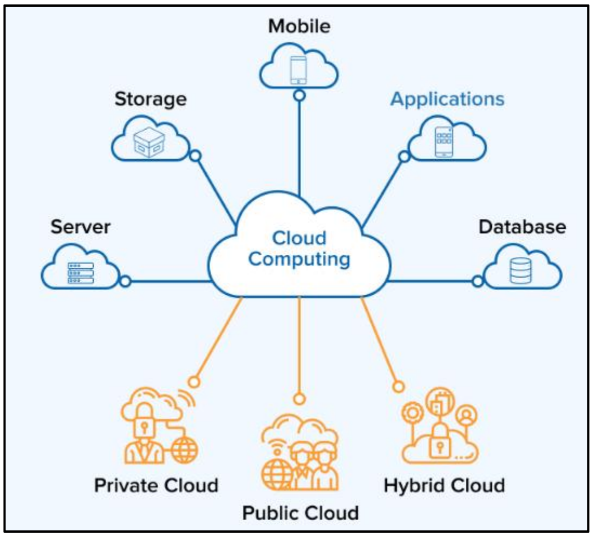


Fig. 2 Services of cloud computing [https://www.tatvasoft.com/blog/cloud-computing-models/]

III. TYPES OF CLOUDS

These are different types of clouds that can subscribe depend upon consumer need such as home user, small business owner, organization and universities need, on base of subscription base, consumer need cloud can be classified into.

- Public Cloud,

- Private Cloud,
- Community Cloud,
- Hybrid or Mixed cloud.

IV. NOVEL CLOUD ARCHITECTURES

In commercial clouds large data centers are operated in a centralized fashion. This design results in high manageability and economy-of-scale. The limitations of this approach are: constructing large data centers such requires huge initial capital investment and high energy expense. Study on size of the datacenters [12] suggests that small size data centres may result in more savings than big data centers in many ways as listed below: 1) less power is consumed by a small data center, hence it does not require a powerful and expensive cooling system. 2) Geographically building the large data centres are expensive than small data centers. 3) Due Geo- diversity of large data center Content delivery and interactive gaming which are time-critical services results in slow performance. For instance, Valances et al. [12-15] presented a feasibility study on hosting of video-streaming services through application gateways. The other related research direction is on using deliberate resources (i.e. resources supplied by users) in deploying cloud applications. If the cloud applications are a mixture of voluntary and dedicated resources then they are much cheaper to operate [15]. These kinds of clouds are more suitable for non-profit applications such as scientific computing. This architecture proposed the new design challenges such as frequent churn events and management of resources that are heterogeneous in nature [16-17].

V. CHALLENGES IN CLOUD COMPUTING

Cloud computing, although slated for an impressive growth, faces numerous challenges. Following is a list of broad categories of these challenges:

- **Security and Compliance:** There are a number of concerns of using cloud for data storage, either as part of an application or as a storage of its own. Many fear about the data security since it is not possible to make physical protection measures (unlike the lock-and-key protections in private premises).
- **Software related Consideration:** Major software related issues are vendor lock-in and interoperability. Many cloud applications are developed to exploit certain cloud specific features and as a result, are not portable.
- **Other Considerations:** Issues such as application complexity and system reliability are also of importance. Most of these considerations would be simplified when the cloud eco-system stabilizes, typically by standardization.

VI. DATA CENTER CONFIGURATION

Total 05 data center have been considered for the simulation environment. Architecture for each data center is given in table 1.

Table 1: Architecture of Data Center

Parameters	**O.S.**
Architecture	X86
Operating System	Linux
Virtual Machine Manager (VMM)	Xen

In our experiment, we have considered the constant pricing for storage and memory for all the data centres. Each Data center will have three hardware units as per the details shown in Table 2.

Table 2: Hardware Units of Data Center

Memory	Storage	Available BW	No. of	Processor
size (Mb)	Details (Mb)		Processor	Speed
2048000	1000000000	10000000	20	10000
2048000	1000000000	10000000	4	10000
2048000	1000000000	10000000	4	10000

DOI: 10.48175/IJARSCT-19647

VII. PROBLEM STATEMENT

When a request arrive to the Datacentre Controller, it has to be allocated to one of the nodes composed the cluster, but the requests have to be distributed evenly and equally among the system to avoid workloads and degradation of system's performance; we need another components to balance the load overall the system called Load Balancer [18].

The Load Balancer plays a very important role in the overall response time of the cloud. In Cloud Computing Scenario Load Balancing is composed of selecting Data Center for upcoming request and Virtual machine management at individual Data Center So, how we can guarantee a good quality of service though balancing the load in the cloud. Our aim is to design a new Load Balancer to improve quality of service by optimizing load balancing in cloud computing.

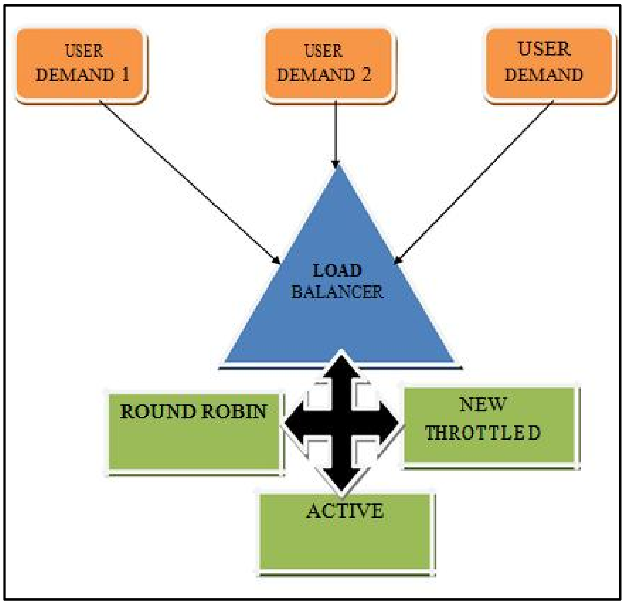


Fig. 3: Load balancing Model

VIII. SOLUTION DOMAIN

Round robin performs the basic type of load balancing and functions simply by providing the list of IP address of cloudlet. It allocates first IP address to the first requester then second IP address to the second requestor for a fixed interval of time known as time slice. If the request is unable to finish within the given slice time, it will have to wait for the next cycle to get it turn for execution. This will continue till submitted tasks are not completed.

IX. IMPLEMENTATION TOOL

A. Cloud Analyst

Cloud Analyst is a GUI based tool that is developed on CloudSim architecture. CloudSim is a toolkit that allows doing modeling, simulation and other experimentation. The main problem with CloudSim is that all the work need to be done programmatically. It allows the user to do repeated simulations with slight change in parameters very easily and quickly. The cloud analyst allows setting location of users that are generating the application and also the location of the data centers. In this various configuration parameters can be set like number of users, number of request generated per user per hour , number of virtual machines, number of processors, amount of storage, network bandwidth and other necessary parameters. Based on the parameters the tool computes the simulation result and shows them in graphical form. The result includes response time, processing time, cost etc. By performing various simulations operation the cloud provider can determine the best way to allocate resources, based on request which data center to be selected and can optimize cost for providing services.

DOI: 10.48175/IJARSCT-19647

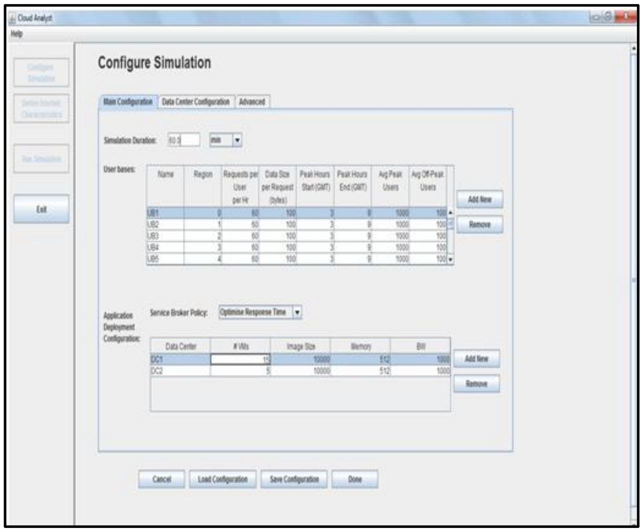
B. Simulation Parameters

Region

In the Cloud Analyst the world is divided in to 6 'Regions' that coincide with the 6 main continents in the World. The other main entities such as User Bases and Data Centers belong to one of these regions. This geographical grouping is used to maintain a level of realistic simplicity for the large scaled simulation being attempted in the Cloud Analyst.

Users

A User Base models a group of users that is considered as a single unit in the simulation and its main responsibility is to generate traffic for the simulation. A single User Base may represent thousands of users but is configured as a single unit and the traffic generated in simultaneous bursts representative of the size of the user base. The modeller may choose to use a User Base to represent a single user, but ideally a User Base should be used to represent a larger number of users for the efficiency of simulation.



Configure Simulation

Simulation Duration: 60.0 min

User bases:

Name	Region	Requests per User per Hr	Data Size per Request (bytes)	Peak Hours Start (GMT)	Peak Hours End (GMT)	Avg Peak Users	Avg Off-Peak Users
UB1	0	60	100	3	9	1000	100
UB2	1	60	100	3	9	1000	100
UB3	2	60	100	3	9	1000	100
UB4	3	60	100	3	9	1000	100
UB5	4	60	100	3	9	1000	100

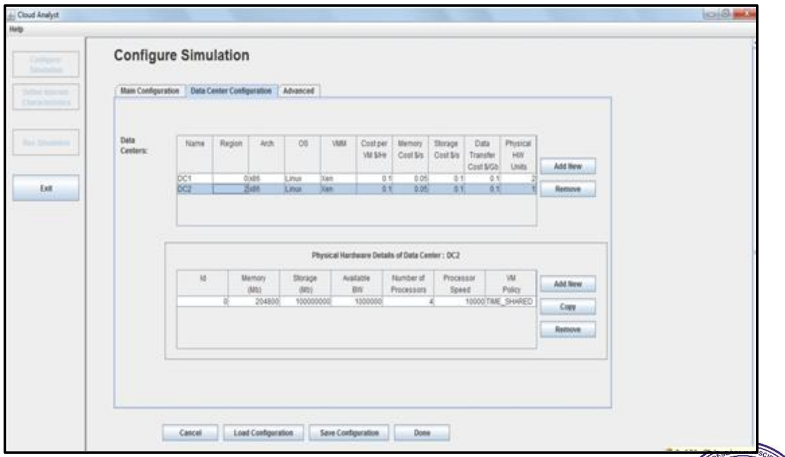
Application Deployment Configuration: Service Broker Policy: Optimise Response Time

Data Center	# VMs	Image Size	Memory	BW
DC1	15	10000	512	1000
DC2	5	10000	512	1000

Fig. 4 configure simulation

C. Data Center Controller

The Data Centre Controller is probably the most important entity in the Cloud Analyst. A single Data Centre Controller is mapped to a single cloudsim. Data Centre object and manages the data centre management activities such as VM creation and destruction and does the routing of user requests received from User Bases via the Internet to the VMs. It can also be viewed as the façade used by Cloud Analyst to access the heart of CloudSim toolkit functionality.



Configure Simulation

Data Centers:

Name	Region	Arch	OS	VMM	Cost per VM $/Hr	Memory Cost $/s	Storage Cost $/s	Data Transfer Cost $/Gb	Physical HW Units
DC1	0	x86	Linux	Xen	0.1	0.05	0.1	0.1	2
DC2	2	x86	Linux	Xen	0.1	0.05	0.1	0.1	1

Physical Hardware Details of Data Center : DC2

Id	Memory (Mb)	Storage (Mb)	Available BW	Number of Processors	Processor Speed	VM Policy
0	204800	100000000	1000000	4	10000	TIME_SHARED

Fig. 5 Data centre configuration

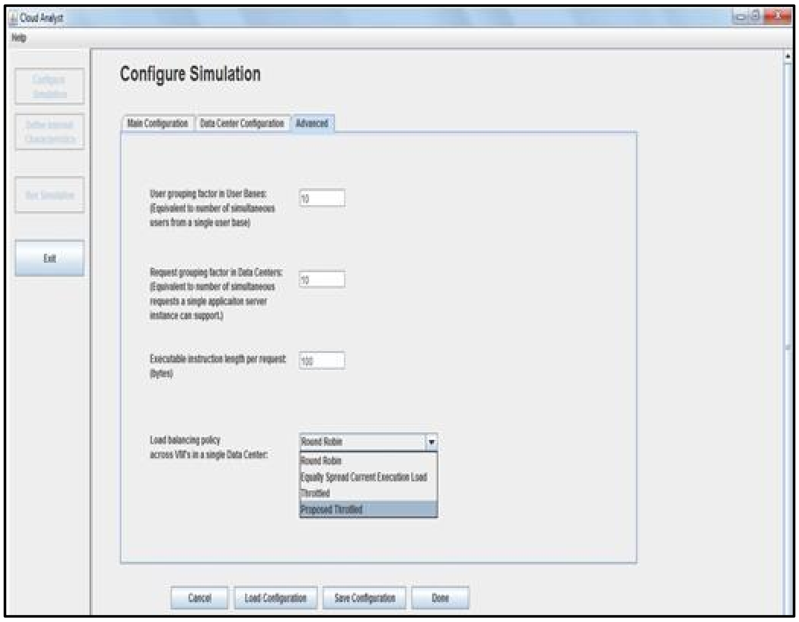
DOI: 10.48175/IJARSCT-19647

Internet Characteristics

In this component various internet characteristics are modeled simulation, which includes the amount of latency and bandwidth need to be assigned between regions, the amount of traffic, and current performance level information for the data centers.

VmLoadBalancer

The responsibility of this component is to allocate the load on various data centres according to the request generated by users. One of the f our given policies can be selected. The given policies are round robin algorithm, equally spread current execution load, throttled, proposed throttled.



Configure Simulation

User grouping factor in User Bases: (Equivalent to number of simultaneous users from a single user base) 10

Request grouping factor in Data Centers: (Equivalent to number of simultaneous requests a single applicaiton server instance can support.) 10

Executable instruction length per request: (bytes) 100

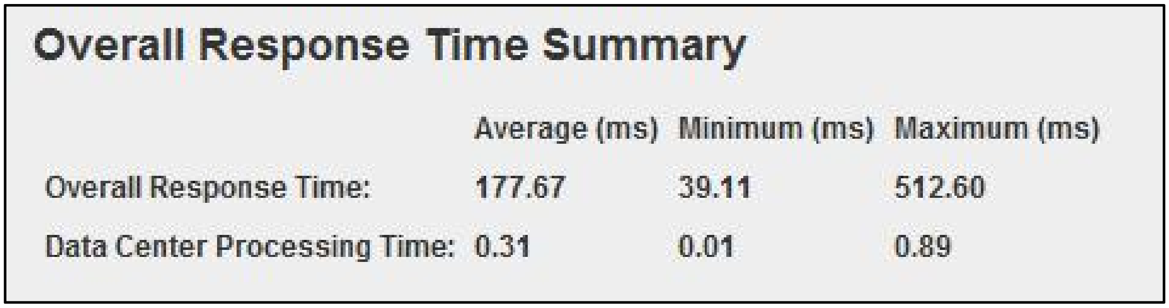
Load balancing policy across VM's in a single Data Center: Round Robin

Fig. 6 Advanced configurations

X. RESULT AND RESPONSE TIME

After performing the simulation the result computed by cloud analyst is as shown in the following figures. The above defined configuration has been used for each load balancing policy one by one and depending on that the result calculated for the metrics like response time, request processing time and cost in fulfilling the request has been shown. Parameters like average response time, data center service time and total cost of different data centres have taken for analysis.

Table: 2 Response Time for RR with 20 User bases

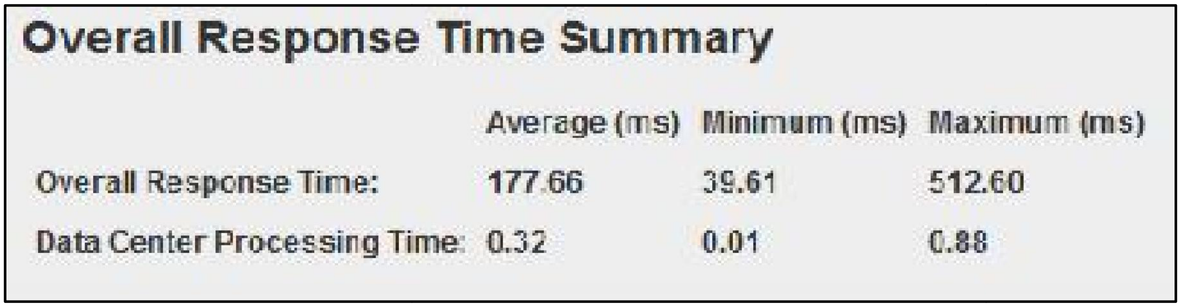


Overall Response Time Summary

	Average (ms)	Minimum (ms)	Maximum (ms)
Overall Response Time:	177.67	39.11	512.60
Data Center Processing Time:	0.31	0.01	0.89

DOI: 10.48175/IJARSCT-19647

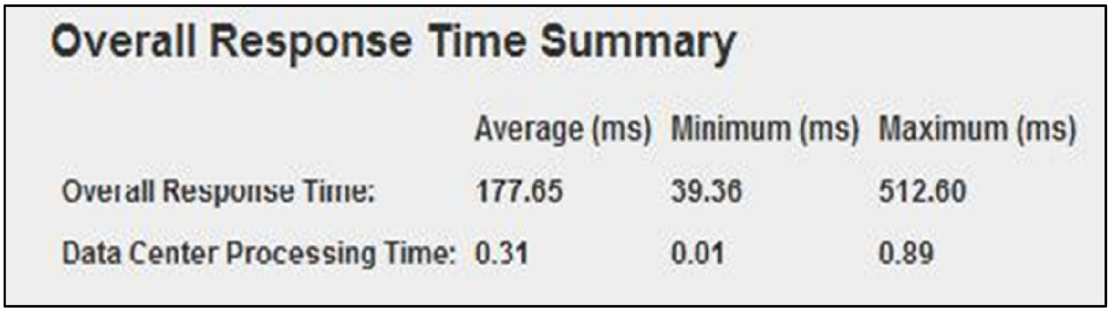
Table: 3: Response Time for ESCE with 6 User bases



Overall Response Time Summary

	Average (ms)	Minimum (ms)	Maximum (ms)
Overall Response Time:	177.66	39.61	512.60
Data Center Processing Time:	0.32	0.01	0.88

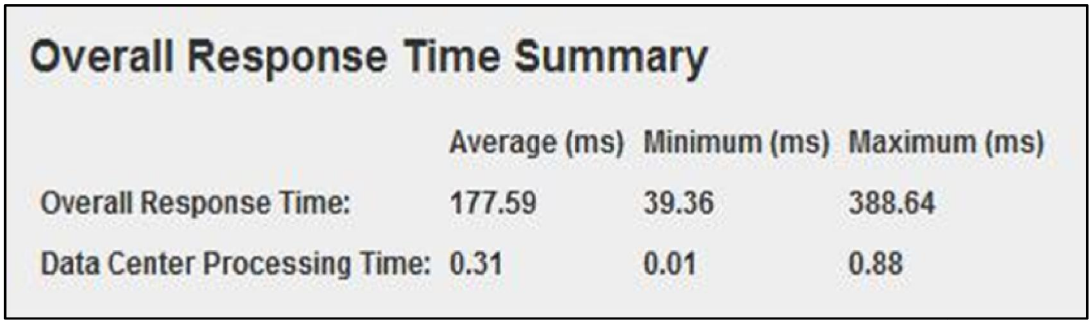
Table 4: Response Time for TLB with 20 User bases



Overall Response Time Summary

	Average (ms)	Minimum (ms)	Maximum (ms)
Overall Response Time:	177.65	39.36	512.60
Data Center Processing Time:	0.31	0.01	0.89

Table 5: Response Time for Proposed TLB with 20 User bases



Overall Response Time Summary

	Average (ms)	Minimum (ms)	Maximum (ms)
Overall Response Time:	177.59	39.36	388.64
Data Center Processing Time:	0.31	0.01	0.88

Key Features of the Proposed Model:

- **Dynamic Load Redistribution:** The model continuously monitors system loads and reallocates tasks in real time, reducing the likelihood of node overloads.
- **Enhanced Throttled Algorithm:** The updated version of the throttled load balancing algorithm improves efficiency by intelligently assigning tasks based on system capacity and resource availability.
- **Comprehensive Algorithm Comparison:** Through Cloud Analyst, we compared multiple algorithms, providing insights into performance metrics such as response time, processing speed, and resource utilization.

Advantages of the Proposed Load Balancing Model:

- **Improved System Availability:** By dynamically balancing the load across nodes, the model ensures that services remain available to clients, even during peak demand periods.
- **Optimized Resource Utilization:** The model maximizes the use of available resources, minimizing idle time and reducing the risk of overloading any single node.
- **Reduced Latency and Response Time:** The updated algorithm enhances system responsiveness, ensuring faster processing of client requests.

- **Scalability:** The proposed model is designed to scale efficiently, accommodating growing workloads in cloud environments without compromising performance.
- **Cost Efficiency:** By avoiding resource over-provisioning and optimizing performance, the model reduces operational costs for cloud service providers.

XI. CONCLUSION

In this work, we have introduced an improved load balancing model tailored for cloud computing environments, specifically presenting an updated version of the throttled load balancing algorithm. Through the use of the Cloud Analyst tool, we conducted a comparative analysis of various load balancing algorithms, highlighting their strengths and limitations. Additionally, we provided a detailed survey of existing load balancing techniques, emphasizing their significance in maintaining efficient cloud operations. In cloud computing, load balancing remains a critical challenge, as it directly impacts the availability and performance of services. The proposed model ensures that client requests are promptly addressed, even under heavy system loads. By redistributing tasks from overloaded nodes to underutilized ones, the load balancer optimizes resource use and prevents performance degradation, ensuring a seamless experience for end users.

REFERENCES

[1]. John Harauz, Lorti M. Kaufinan. Bruce Potter, "Data Security in the World of Cloud Computing", IEEE Security & Privacy, Copublished by the IEEE Computer and Reliability Societies, July/August 2009.

[2]. Patel, K., & Gupta, R. (2021). Geo-diversity and its impact on the performance of content delivery and time-critical services in cloud computing. Future Generation Computer Systems, 125, 77-89. https://doi.org/10.1016/j.future.2021.10.012

[3]. National Institute of Standards and Technology- Computer Security Resource Center - www.csrc.nist.gov

[4]. Samerjeetkaur, "Cryptography and Encryption in Cloud Computing", VSRD International Journal of Computer Science and Information Technology, VSRDIJCSIT, Vol. 2 (3), 2012.

[5]. Patel., S.(2021). Enhancing Image Quality in Wireless Transmission through Compression and De-noising Filters. In International Journal of Trend in Scientific Research and Development (Vol. 5, Number 3, pp. 1318–1323). IJTSRD. https://doi.org/10.5281/zenodo.11195294

[6]. Sun, Y., & Tang, H. (2020). Voluntary resource utilization in hybrid cloud architectures: Design challenges and economic benefits for scientific computing. IEEE Cloud Computing, 7(5), 35-45. https://doi.org/10.1109/MCC.2020.3012504

[7]. Chen, L., & Zhang, J. (2024). Cost-effectiveness of small versus large data centers in cloud computing: A new perspective on resource management. Journal of Cloud Computing, 13(1), 33-48. https://doi.org/10.1186/s13677-024-00211-9

[8]. Ramgovind S, Eloff MM, Smith E, 'The management of security in cloud computing", IEEE – 2010.

[9]. Aderemi A. Atayero and OluwaseyiFeyisetan," Security Issues in Cloud Computing: The Potentials of Homomorphic Encryption" Journal of Emerging Trends in Computing and Information Sciences, VOL. 2, NO. 10, October 2011.

[10]. Turban, E; King, D; Lee, J; Viehland," Chapter 19: Building E-Commerce Applications and Infrastructure". Electronic Commerce A Managerial Perspective.pp. 27, 2008.

[11]. J. Kruskall and M. Liberman."The Symmetric Time Warping Problem: From Continuous to Discrete. In Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison", pp. 125-161, Addison-Wesley Publishing Co., 1983.

[12]. Mr. Nitin S. More, Mrs. Swapnaja R. Hiray and Mrs. SmitaShukla Patel," Load Balancing and Resource Monitoring in Cloud", International Journal of Advances in Computing and Information Researches ISSN: 22774068, Volume 1– No.2, April 2012.

[13]. Patel., S."Optimizing Wiring Harness Minimization through Integration of Internet of Vehicles (IOV) and Internet of Things (IoT) with ESP-32 Module: A Schematic Circuit Approach", International Journal of

Science & Engineering Development Research (www.ijrti.org), ISSN:2455-2631, Vol.8, Issue 9, page no.95 - 103, September-2023, Available :http://www.ijrti.org/papers/IJRTI2309015.pdf

[14]. R. X. T. and X. F. Z,"A Load Balancing Strategy Based on the Combination of Static and Dynamic, in Database Technology and Applications (DBTA)",2nd International Workshop,2010.

[15]. Patidar, M. et al. (2024). Efficient Design of Half-Adders and EXOR Gates for Energy-Efficient Quantum Computing with Delay Analysis Using Quantum-Dot Cellular Automata Technology. In: Al-Turjman, F. (eds) The Smart IoT Blueprint: Engineering a Connected Future. AIoTSS 2024. Advances in Science, Technology & Innovation. Springer, Cham. https://doi.org/10.1007/978-3-031-63103-0_22

[16]. Patel., S.(2022). Performance Analysis of Acoustic Echo Cancellation using Adaptive Filter Algorithms with Rician Fading Channel. In International Journal of Trend in Scientific Research and Development (Vol. 6, Number 2, pp. 1541–1547). IJTSRD. https://doi.org/10.5281/zenodo.11195267

[17]. Zhang, L., Liu, X., & Wu, Y. (2023). Energy efficiency and cooling optimization in small-scale data centers: A comparative analysis with large data centers. IEEE Transactions on Cloud Computing, 11(2), 145-158. https://doi.org/10.1109/TCC.2023.3124567

[18]. Valances, M., & Smith, A. (2022). Feasibility study on video-streaming services through application gateways: A case study on data center architecture. Journal of Cloud Applications, 14(3), 221-235. https://doi.org/10.1016/j.jca.2022.05.003