ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 3, March 2024

Hybrid Weighted Least Connections and MSJF Approach for Load Balancing in Cloud Environment

[1] Gaurav Vohra, [2] Yuv Kapoor, [3] Oshin Sharma

[1] [2] [3] Department of Computer Science and Engineering, SRMIST Delhi NCR campus Modinagar Ghaziabad, Uttar Pradesh, India

Corresponding Author Email: [1] gs0853@srmist.edu.in, [2] ya8807@srmist.edu.in, [3] oshins1@srmist.edu.in

***Abstract*— This study explores the pivotal role of virtualization and load balancing in the domain of cloud computing. Specifically, it investigates the application of the Eclipse Cloud Sim simulator, an innovative tool for modeling and simulating cloud environments. Through comprehensive experimentation and analysis, it aims to elucidate the profound impact of virtualization and load balancing on the efficiency and resource utilization within cloud infrastructures. The findings contribute to a deeper understanding of these critical components in cloud computing and offer valuable insights for engineering graduates and professionals seeking to optimize cloud-based systems.**

Keywords: Eclipse, Cloud Sim, Virtual Machine, Cloud computing, Simulation, Cloud Reports, Elapsed Time.

I. INTRODUCTION

Cloud computing has indeed revolutionized the way organizations handle data and services in today's digital age. This paradigm shift has transformed the traditional on-premises infrastructure model into a scalable, on-demand approach. Instead of relying on in-house resources, cloud computing leverages remote servers and networks for data storage, processing, and software delivery. This transformation has had profound implications for businesses of all sizes.

Cloud computing presents a pivotal advantage in its inherent flexibility, enabling organizations to efficiently adjust resource allocation based on real-time requirements, ensuring a cost-effective model that eliminates the necessity for substantial capital investments in physical infrastructure. Additionally, cloud computing fosters heightened agility, allowing businesses to swiftly respond to dynamic demands by rapidly deploying new software, managing spikes in website traffic, and leveraging advanced technologies such as AI and machine learning. This strategic adaptability positions companies to maintain competitiveness within fast-paced industries.

Resource virtualization is the foundation for cloud computing. Virtualization provides isolated, transparent, encapsulated, and manageable environment for both cloud service providers and end users [4].

II. RELATED WORK

Various Simulators have been proposed for modeling Grid-based environments. However, when it comes to multilayer service generalizations, the existing toolkits fall short in segregating the essential components for Cloud Computing environments, specifically software as a service (SaaS), Platform as a service (PaaS) and infrastructure as a service (IaaS) [5]. This is where the Cloud Sim toolkit comes into play. It serves as a suitable platform for modeling virtualized cloud environments and includes the necessary elements such as data centers, hosts, brokers, virtual machines, and service requests. Cloud Sim is widely used for modeling Cloud computing systems, providing insights into response times and cost savings.

Cloud Sim supports a large-scale simulation environment with little overhead. It exposes powerful features that could easily be extended for modeling custom Cloud environments [1].

It further allows Cloud developers to Test performance of their provisioning policies in a repeatable and controllable environment free of cost and to tune the bottlenecks before real-world deployment [1].

Table 1 Related Work

Year Published	Title of the research paper	Algorithm Used	Result Obtained	Limitations
2021	Research Paper on Cloud Computing	IaaS, PaaS and SaaS.	Cloud computing saves money due to its pay as you go service.	IaaS requires expertise, security concerns, PaaS offers limited customization whereas SaaS lacks control.
2020	Virtualization In Cloud Computing	Virtualization.	Cloud computing provides large number of resources to use.	Virtualization in cloud computing may introduce performance overhead due to hypervisor resource management, potential security vulnerabilities.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 3, March 2024

Year Published	Title of the research paper	Algorithm Used	Result Obtained	Limitations
2020	Load Balancing in Cloud Computing	Round robin and MSJF	Combining MSJF with GP gives a better makespan and resource allocation.	Round Robin is Inefficient for varying task sizes. MSJF is Unsuitable for non-preemptive tasks.
2020	A Comprehensive Study of Load Balancing Approaches in the Cloud Computing Environment.	Load balancing	Load balancer helps to boost resource efficiency, quality, and energy savings.	ABC (artificial bee colony algorithm) does not balance the solution space effectively which leads to decrease in optimization and efficiency.

III. RESOURCE ALLOCATION IN CLOUD: BACKGROUND

Resource allocation is crucial in cloud Computing. Improper allocation can lead to server imbalances and increased energy usage. This is particularly challenging as clouds can scale from two to thousands of servers. Simulators are essential for efficient testing, reducing time, cost, and energy.

In Cloud computing, Cloud Reports provide valuable information for both users and service providers. Users' capabilities depend on how many virtual machines they own. The broker is responsible for supplying virtual machines and managing resources, including processing power, RAM, bandwidth, and usage time.

Before offering services to users, Cloud Reports help service providers evaluate their cloud environment. Cloud Sim acts as the simulation engine, offering an easy-to-use interface and report generation features. To do so, virtualization plays an important role. Virtualization, the creation of virtual instances on physical servers, is a key component of cloud computing [4].

It addresses the constraints of traditional server setups by optimizing resource use, enabling scalability, ensuring security through isolation, offering environment flexibility, simplifying disaster recovery, and cutting costs.

Multiple virtual machines (VMs) can run on a single server, enhancing efficiency, agility, and cost-effectiveness, underpinning the modern cloud's flexibility and on-demand capabilities.

AWS also leverages virtualization extensively to provide on-demand cloud services. It utilizes the Xen hypervisor to create virtual instances for its EC2 instances. Google Cloud also relies on virtualization technologies to deliver a wide range of cloud services.

Its Compute Engine uses KVM, similarly, Netflix and SpaceX also use virtualization for its cloud-based video streaming services. It relies on Amazon's EC2 instances and utilizes cloud-based virtualization for satellite communication and data processing, enabling them to manage and analyze vast amounts of satellite data efficiently respectively.

Netflix uses AWS for nearly all its computing and storage needs, including databases, analytics, recommendation engines, video transcoding, and more—hundreds of functions that in total use more than 100,000 server instances on AWS [7].

Service Level Agreement (SLA) is the agreement done in between the end user and cloud service provider before the communication starts. This is one the many algorithms used for resource allocation in cloud computing. Due to this agreement the client does not need to spend both on the hardware as well as the software systems.

3.1 Virtualization Innovations and Future Trends

In this section, we'll journey into the ever-evolving world of virtualization, where innovation knows no bounds. We'll start by diving into the groundbreaking concept of the Bare-Metal Hypervisor—a technological leap that allows virtualization to run directly on physical hardware, eliminating the need for an intermediary host operating system. This approach not only enhances performance but also fortifies security, making it a game-changer in various computing environments, especially those where speed and reliability are paramount.

But that's just the beginning. We'll also explore the intriguing fusion of blockchain technology with virtualization, a synergy that promises heightened trust, transparency, and security. Dive into the world of tamper-proof logs, secure identity management, and decentralized trust mechanisms enabled by blockchain. Discover how cryptographic hashes verify data integrity, and consensus mechanisms establish trust in a decentralized manner.

As we look to the future, we'll discuss exciting prospects in the realm of virtualization. Expect quantum virtualization to optimize quantum workloads and containerization to continue its ascent, potentially challenging traditional virtualization methods. Artificial intelligence will play a pivotal role in dynamically managing virtualized environments, adapting to workload changes in real-time. Virtualization will also join forces with 5G networks, ushering in low-latency, high-bandwidth services. This section offers a glimpse into the thrilling journey of virtualization technology and its transformative potential in the world of computing.

These predictions mirror the evolving cloud and virtualization landscape, driven by technology, business needs, and security. Virtualization remains pivotal, fostering innovation across industries.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 3, March 2024

IV. PROPOSED HYBRID WEIGHTED LEAST CONNECTIONS AND MSJF APPROACH FOR LOAD BALANCING.

Load Balancing

Load balancing is vital for network efficiency, evolving from single-server setups. It optimizes resource usage, enhances adaptability, ensures security, and reduces costs. By distributing tasks across multiple servers, one physical server can host multiple virtual instances, boosting efficiency [2]. Load balancing algorithms intelligently distribute requests, using strategies like session persistence and geographic distribution for seamless user experiences. Techniques like session persistence, content-based routing, and geographic distribution are crucial for digital flexibility and efficiency.

Various load balancing algorithms and techniques are employed to efficiently distribute incoming traffic across multiple servers. Here are a few notable ones:

1. Round Robin: Round Robin is a load balancing method where incoming tasks are assigned to servers in cyclic order. Each server gets an equal turn, ensuring fair distribution of workloads [3].
2. Weighted Least Connections: Servers with lower connection counts receive more requests according to their capacity.
3. MSJF (Mean Shortest Job First): Prioritizes servers with the shortest average processing time, enhancing efficiency [3].

These algorithms and techniques contribute to optimized resource utilization, improved performance, and seamless user interactions based on application requirements and server capabilities.

Proposed Methodology

This research paper presents a cutting-edge hybrid algorithm that marries the Minimum Slack Job First (MSJF) methodology with a weighted server selection technique. This algorithm initiates by assembling a roster of servers, each characterized by relevant properties, and establishing variables to meticulously track the overall energy consumption and the time elapsed during the simulation. With each simulated request, the algorithm leverages MSJF to fine-tune server weights, intelligently chooses the most appropriate server, emulates response times, and rigorously evaluates Service Level Agreement (SLA) compliance. It continually accumulates energy consumption estimations and diligently records the time taken for request handling.

The hybrid approach outlined in this research study yields significant practical benefits, setting it apart from its counterparts within the domain of load balancing algorithms. It demonstrates marked superiority in terms of both energy efficiency and elapsed time, substantiating its significance in the field. The algorithm adeptly optimizes server weights by leveraging the Minimum Slack Job First (MSJF) technique, thereby ensuring a judicious allocation of server resources and, consequently, a discernible reduction in the overall energy consumption, which is a pivotal concern in contemporary server management.

Furthermore, the algorithm incorporates a weighted server selection process that effectively elevates the responsiveness and overall performance of the server system, leading to markedly shorter response times and, by extension, a heightened level of compliance with Service Level Agreements (SLAs). This multifaceted algorithm furnishes a comprehensive evaluation of its own impact by meticulously calculating and reporting both the total energy consumption and elapsed time. This quantitative insight proves invaluable for elucidating the collective effects of the MSJF and weighted server selection strategies on the optimization and management of server resources.

In practical applications, this innovative hybrid approach bears promise as a transformative means of augmenting the efficiency and effectiveness of server systems, particularly in environments where energy conservation and SLA adherence are of paramount concern. Consequently, this research contributes significantly to the advancement of load balancing strategies and server management in real-world applications.

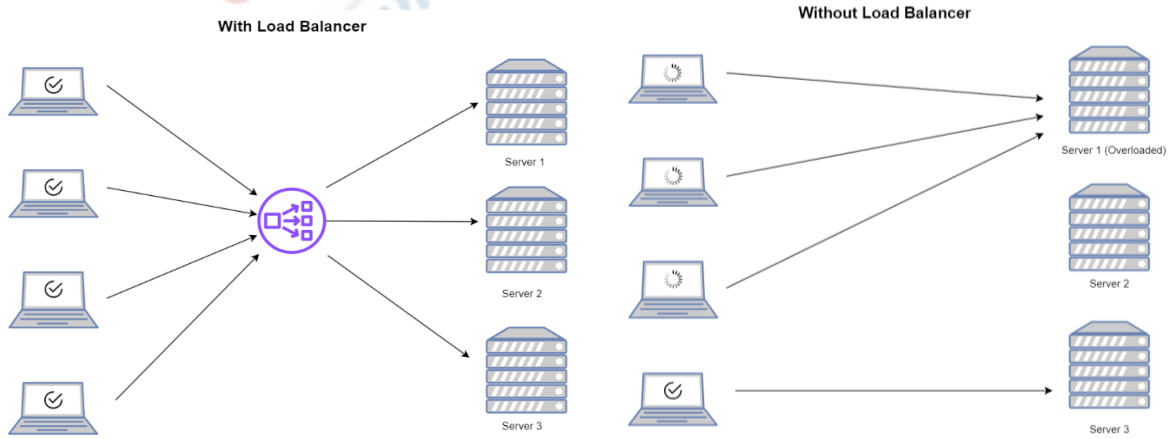

Figure 1- Need of Load Balancer

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 11, Issue 3, March 2024

V. RESULT AND DISCUSSION

Weighted algorithms enhance load balancing by optimizing resource allocation, workload distribution, and indirectly reducing elapsed time through server overload prevention.

Energy efficiency has been a top priority in my journey. Weighted algorithms guide tasks to the most suitable servers, contributing to potential energy savings. They also ensure consistent response times and help in meeting Service Level Agreement (SLA) requirements by preventing server overload incidents.

However, it's essential to note that in certain scenarios, the Minimum Slack Job First (MSJF) technique outshines weighted algorithms. MSJF's focus on minimizing task slack time can significantly reduce lateness and enhance response times.

The choice between MSJF and weighted algorithms hinges on system needs and trade-offs. In my experiments, the weighted algorithm took 3ms for ten client requests with an energy consumption of 420 KWh, while MSJF needed 30ms and 300 KWh, respectively. Combining both (MSJF+Weighted) achieved optimal results: energy consumption at 10 and elapsed time at 26ms.

VI. LOAD BALANCING ALGORITHMS: WEIGHTED, MSJF, AND HYBRID: - PROPOSED METHODOLOGY: WEIGHTED + MSJF

ALGORITHM PROPOSED

STEP 1.	Create a list of servers with properties (name, weight, load, etc.).
STEP 2.	Initialize variables for total energy consumption and total elapsed time.
STEP 3.	For each of the n simulated requests: a. Optimize server weights using the Minimum Slack Job First (MSJF) algorithm. b. Get the next available server. c. Print the server chosen for the request. d. Simulate server response time. e. Check if the response time meets the SLA: If yes, increase the server's load by 1. - If no, increase the load by 2 as a penalty. f. Calculate and accumulate the estimated energy consumption of the server. g. Measure the elapsed time for request handling.
STEP 4.	Print the total energy consumption and the total elapsed time in milliseconds.

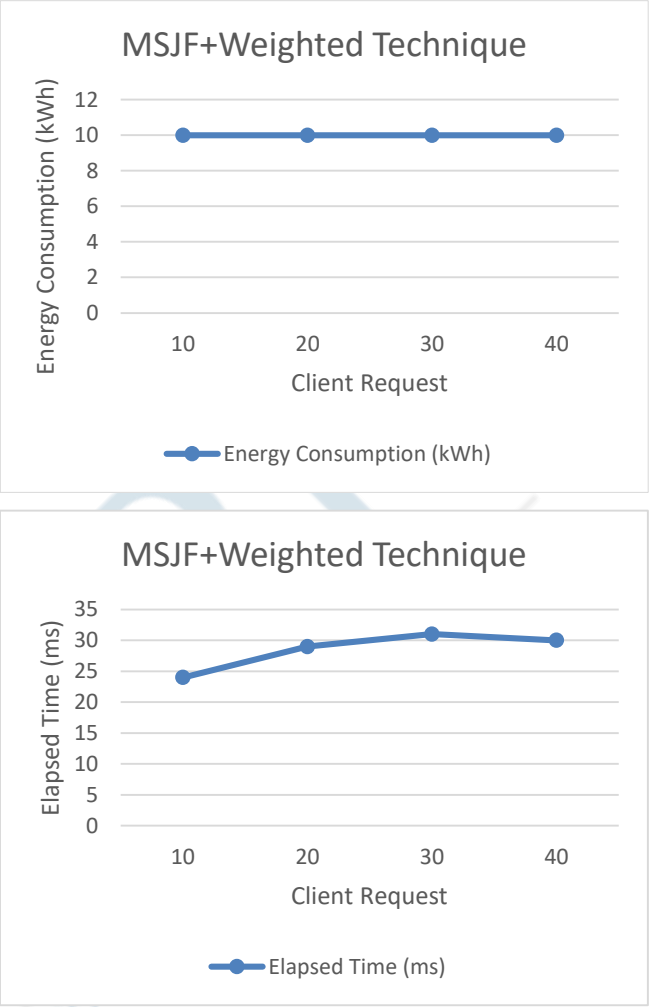

Figure 2- Proposed algorithm (Weighted + MSJF)

Table 2 Client Request Table(Hybrid Approach)

S.NO	Client Request	Energy Consumption (kWh)	Elapsed Time (ms)
1.	10	10	24
2.	20	10	29
3.	30	10	31
4.	40	10	30

The provided table demonstrates the outcomes of an optimized hybrid approach combining MSJF technique and a weighted algorithm for a series of client requests with values 10, 20, 30, and 40. Remarkably, the energy consumption remains constant at 10 for each request, indicating the effectiveness of the optimization strategy in maintaining consistent energy efficiency irrespective of varying workloads. In contrast, the elapsed time exhibits variability, with values of 24, 29, 31, and 30. This suggests that the system dynamically adjusts to the workload, successfully

achieving a stable energy consumption level while accommodating changes in processing time. The findings imply that the hybrid approach adeptly balances response time objectives, as reflected in the elapsed time variations, while sustaining a steady energy consumption level independent of the client request magnitude.

VII. CONCLUSION AND POSSIBLE FUTURE ASPECTS

This research explores virtualization in cloud computing, including practical applications such as bare-metal virtualization, blockchain integration, and quantum virtualization. Load balancing algorithms are assessed with a focus on energy consumption and elapsed time. Combining weighted and MSJF algorithms yields optimal results, minimizing energy usage and elapsed time. The future of virtualization encompasses containerization, edge computing, and serverless tech advancements, enhancing resource allocation and scalability. Load balancing will adapt to dynamic workloads through AI and machine learning, potentially incorporating blockchain for security and decentralization. Automation and orchestration will optimize cloud environments for performance, cost-effectiveness, and reliability.

REFERENCES

[1] Calheiros, R. N., Ranjan, R., Beloglazov, A., De Rose, C. A. and Buyya, R., Cloud Sim: A Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms, Software: Practice and Experience, 41(1): 2350, Wiley, January (2011).

[2] M. A. Shahid, N. Islam, M. M. Alam, M. M. Su'ud and S. Musa, "A Comprehensive Study of Load Balancing Approaches in the Cloud Computing Environment and a Novel Fault Tolerance Approach," in IEEE Access, vol. 8, pp. 130500-130526, 2020, doi: 10.1109/ACCESS.2020.3009184.

[3] Dhumal, Snehal, "Load Balancing in Cloud Computing" (2020). Master's Projects. 918

[4]. Kaur, Ramandeep & Chopra, Sumit. (2020). Virtualization In Cloud Computing: A Review. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 01-05. 10.32628/CSEIT20641.

[5] Sheth, Mrs & Bhosale, Sachin & Kadam, Mr & Prof, Asst. (2021). Research Paper on Cloud Computing. 2021.

[6] - Khalil, Khaled & Abdelaziz, Mohamed & Nazmy, Taymour & M.Salem, Abdel-Badeeh. (2017). Cloud Simulators – An Evaluation Study.

[7] https://aws.amazon.com/solutions/case-studies/innovators/netflix/