Advancements in Green Cloud Computing: A Survey of Resource Management and Carbon Emission Reduction

M. Dhanalakshmi[1], Dr. S. Manoharan[2]

[1]Research Scholar, Department of Computer Applications, Nandha Arts and Science College (Autonomous), Erode - 638 052.
Email: *dhanalakshmibca3[at]gmail.com*

[2] Principal and Professor of Computer Science, Nandha Arts and Science College (Autonomous), Erode 638 052

Abstract: ***Green Cloud Computing has become a sustainable computing paradigm which seeks to minimize energy usage, carbon dioxide emissions and the cost of operating systems while keeping performance peak in Cloud Computing systems with a high level of services. Clustering of all cloud computing, big data analytics, Artificial Intelligence (AI) is putting enormous pressure on the world's economy and the environment to supply the energy needed for its escalating growth. Efficient use of the computer resources such as job scheduling, load balancing and fault tolerance have become essential to maximize cloud operations and ensure sustainable utilization of computer resources. Task scheduling allows to schedule tasks to the right resources to minimize execution time and energy, load balancing allows to distribute the tasks equally between servers to avoid hot spots and optimize the use of resources. To ensure service availability and system reliability, fault-tolerance mechanisms offer fault detection, prediction and recovery in the event of hardware, software and network failure. To solve these challenges, the Deep Learning (DL) techniques have been found to be effective solutions for the intelligent workload prediction, adaptive resource allocation, proactive fault detection and automatic decision making. Cloud systems can be designed to be environmentally friendly while providing good service quality and performance by integrating carbon emission factors with scheduling, load balancing and fault-tolerant decision making. Combining DL models with the job scheduling, load balancing and fault-tolerance features of the cloud computing systems, cloud service providers can optimize even further energy consumption, lower carbon footprint, enhance Quality of Service (QoS) and improve service dependability. Optimization algorithms, Reinforcement Learning (RL), model-based methods, Neural Networks and hybrid intelligent approaches have been the recent developments with great promise for developing sustainable and resilient cloud computing environments. This article discusses several cloud-based research efforts and offers a thorough overview and comparison of current strategies for task scheduling that is both energy efficient and aware of carbon emissions. By systematically reviewing existing resource management techniques, this study highlights critical gaps and opportunities for improvement. Researchers interested in innovative cloud computing and discovering ways to lower emissions of carbon in cloud settings will find the results useful.***

Keywords: Green Cloud Computing, Task Scheduling, Load Balancing, Fault-Tolerance, Carbon Emission, Resource Utilization

1. Introduction

The term "cloud computing" refers to the practice of renting out various Information Technology (IT) resources like as servers, networks, databases, storage, applications, and more over the internet on an as-needed basis. The term "cloud computing" refers to a method of storing and retrieving data and programs over the internet as opposed to a traditional hard drive. Among the many types of data are audio and video recordings, documents, images, and files. In addition, with cloud computing, users may access virtual resources from anywhere with an internet connection [1].

The cloud computing platform can be broadly divided into service models and deployment models. The management and delivery of cloud resources to users are established through the usage of cloud deployment models. Three primary types are available for deployment:

- Public clouds are services accessed through the internet and managed by a third party, which are suitable for large scale applications.
- Private clouds are for a single organization and provide greater security and control. To optimize resource use and be flexible,
- Hybrid clouds merge public and private cloud environments. Community clouds are those that are accessed by organizations that have similar goals and policies or related security needs. [2]

On a subscription-based, pay-as-you-go basis, users have access platform, software (applications), and infrastructure through a cloud system. These services are known by different names in the industry: Infrastructure as a Service (IaaS), which refers to the practice of outsourced hardware, processors, memory, and networking components to perform this operation; Platform as a Service (PaaS), which provides developers with a group of software parts in order to develop applications; and Software as a Service (SaaS), which makes end-user programmers available online and runs on a shared set of hardware infrastructure [3].

The focus of green computing research is on creating effective clouds with eco-friendly features like energy monitoring, virtualisation, high-efficiency computing, burden balancing, green data centres, extensibility and recycling, among others [4]. This is designed to track the overall energy consumption associated with user requests. Its principal parts are the Carbon Emission Directories and the Green Cloud offering, which monitor the energy efficiency of every cloud provider and give them incentives to green their services. From the user's point of view, the Green Broker plays an important role for monitoring and selecting cloud services that meet their QoS requirements while ensuring the lowest possible carbon emissions during service provision [5]

The goal of cloud task scheduling is to reduce execution time by allocating computer resources to jobs in the most efficient manner or by choosing the best suitable resource for task execution. Prioritizing work based on a number of criteria is the first step in scheduling algorithms, enabling them to generate a list of assignments. After then, jobs are given priority and allocated to processors and computers that can complete them according to a set of predefined criteria. Figure 1. Shows the Task Scheduling

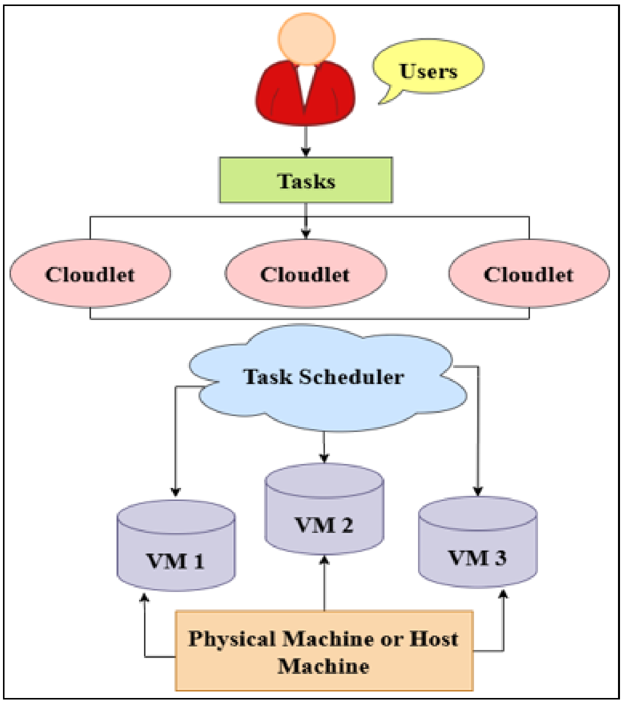

Figure 1: Task Scheduler

- Static scheduling is the process of organizing jobs within a defined environment, which encompasses the structure of tasks, allocation of resources and projected execution durations.
- Dynamic scheduling in cloud environments takes into account both the tasks that have been submitted and the current states of the system and machines. This approach ensures optimal decision-making regarding how tasks are scheduled based on real-time conditions. [6]

If there is an excess of dynamic local workload in the cloud, load balancing can be used to disperse it evenly among all the nodes. Improving the overall efficiency of the system is achieved by avoiding overloading any one node, which leads to good user experience and resource usage ratio. An effective method of maximizing available resources is load balancing, which reduces resource use. It leads to better utilization of resources, system performance, faster response times and increased user satisfaction. [7] In green cloud setups, load balancing plays a vital role in energy efficiency by distributing tasks among servers evenly, avoiding resource hotspots and minimizing power usage. Efficient load balancing can also reduce the waste of resources, promote scalability, fault-tolerance and service reliability. There is a close connection between the amount of energy used and the amount of carbon emissions generated, so using energy wisely also indirectly reduces the carbon footprint of cloud data centers. Therefore, load balancing is seen as a crucial strategy for environmentally friendly and energy-efficient cloud computing settings. The Load Balancing is shown in Figure 8.

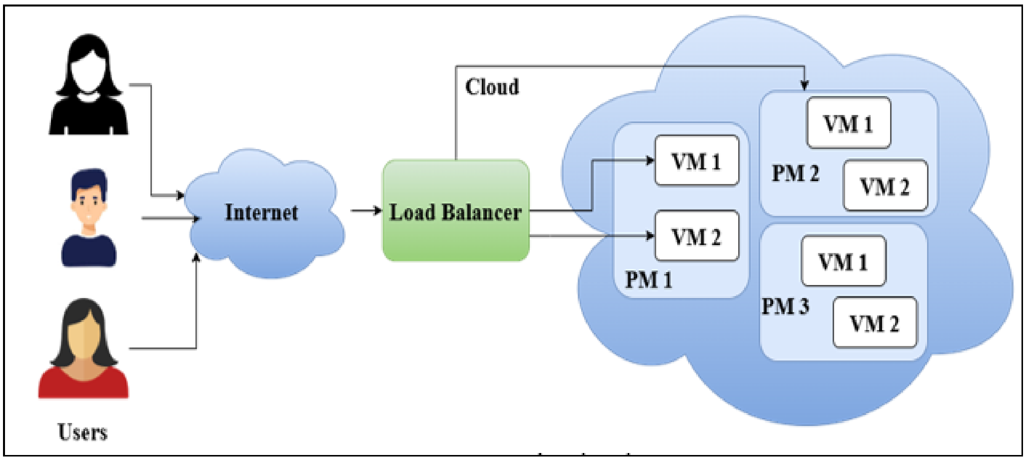

Figure 2: Load Balancing

Fault tolerance is an important mechanism in Green Cloud Computing, which allows the cloud system to maintain normal operation even if there are faults or failures. It increases the reliability, availability and dependability of cloud services by identifying and recovering from faults before they impact users. In green cloud systems, fault tolerance can help ensure service continuity, optimize resource utilization, and save energy use. To ensure that applications can keep running in the modern era of energy-efficient and environmentally friendly cloud computing, robust fault-tolerance methods are essential. Figure 3 shows the five different kinds of faults that can occur in green cloud computing.

- Physical Faults (Hardware Faults): Faults that relate to hardware components like CPU failure, memory fault,

Paper ID: SR26827150359 DOI: https://dx.doi.org/10.21275/SR26827150359

storage device failure, overheating and power supply failure.
- Network Faults (Link Faults): These faults are attributed to the communication faults in the network with the occurrence of packet loss, failure of links, network congestion and connectivity that prevents access to cloud resources.
- Processor Faults (Node Faults): These faults are caused by software bugs, insufficient resources, inefficient processors, operating system problems or failed computing nodes.
- Service Expiry Faults: These are faults that are caused due to the expiry of the allocated service time or resource lease during the time that the application is still using the resource, which leads to the service disruption.
- Timing Faults: These faults happen when an application exceeds the specified time limit, resulting in poor performance and QoS violations. [9,10]

A green cloud computing system that is both energy efficient and sustainable was achieved through the application of DL models for work scheduling, load balancing, and fault-tolerance mechanisms. Intelligent task scheduling optimizes resource allocation to save energy, while load balancing prevents resource overloading and performance degradation. The aim of fault-tolerance strategies is to maintain the reliability of the frameworks by detecting faults and mitigating their impact. These technologies improve energy efficiency, reduce costs, enhance QoS and promote environmentally friendly cloud data centers.

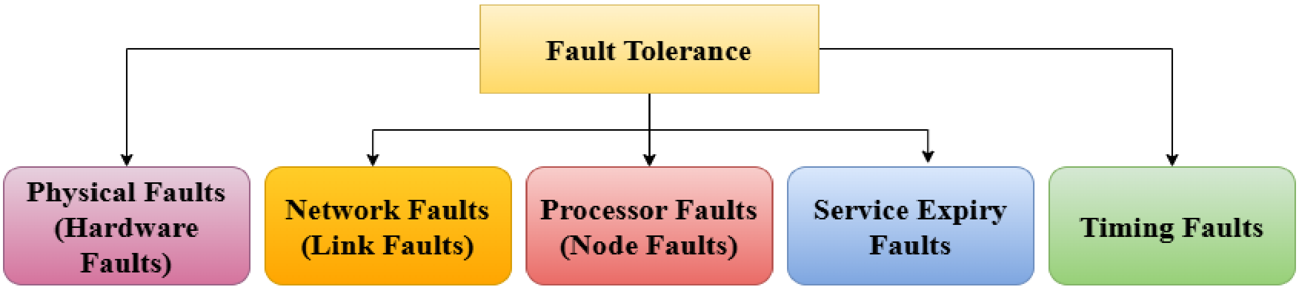

Figure 3: Fault-Tolerance

Examples from Green Cloud-based Task Scheduling, Load Balancing, and Fault-tolerance show the ongoing innovations in this domain, which are described in this study. Task scheduling, load balancing, and fault tolerance based on the green cloud are introduced in Section 2. Section 3 presents the contrasted evaluation of the examined models. The methods' safety and efficacy will be assessed in Section 4. Section 6 wraps up the study, while Section 5 lays out potential avenues for further research.

2. Literature Survey on Green Cloud based Task Scheduling, Load Balancing and Fault-Tolerance

In order to support the concept of green computing, Ari et al. [11] used a modified ant colony optimization-based biologically inspired scheduling approach to distribute workloads evenly across available resources while keeping makespan time to a minimum. In order to solve the work scheduling problem optimally, Cloud- Ant Colony Optimization (CACO) mostly uses ACO capabilities. However, cloud computing does not work well with conventional job scheduling techniques.

To efficiently run task-based applications on shared computing platforms and save energy consumption, Juarez et al. [12] introduced a real-world dynamic scheduling system. The multi-heuristic resource allocation (MHRA) approach is used for partial solutions; it is basically a faster local search strategy. The program aims to optimize energy consumption and execution time while taking into consideration sequence-dependent setup durations, VM setup, downtime, and energy profiles that are specific to different architectures. A prototype scheduler was tested on multiple random Directed Acyclic Graphs (DAGs) and real task-based COMPSs applications. However, this model highlights a trade-off between energy efficiency and execution speed, indicating that enhancing performance may lead to increased energy usage.

In a geographically shared heterogeneous hybrid cloud environment, Zhao et al. [13] created a work scheduling system that uses PPO to harvest renewable energy. The DRL method match the renewable energy supply with demand through automated task shifting and cloud-bursting during resource shortages. Scheduling and decision-making based on DRL algorithm and Neural Networks with focus on convergence and metrics. But infrastructure changes require model retraining and unexpected workloads can cause a degradation of the policy.

The Improved Chaotic Binary Grey Wolf Optimization (IGWO) method, proposed by Mohammadzadeh et al. [14], is a green cloud computing workflow scheduling tool. This method employed a GWO algorithm by integrating the hill-climbing and chaos theory. Because of this, the GWO was able to keep from falling into a local optimum and achieve faster convergence. The algorithm has been implemented as a binary program by combining the many S and also V functions. Despite that, the computational complexity of this model is larger.

The Savitzky-Golay Long Short-Term Memory networks (SG-LSTM) prediction model was presented by Bi et al., [15] and is used for filtering noise and green energy prediction. Efficiently managing energy costs and revenue in Green Cloud Data Centers (GCDCs) is made easier with a bi-objective optimization approach called Decomposition-based

Multi-objective evolutionary algorithm with Gaussian mutation and Crowding distance (DMGC). This technology can be used for both task scheduling and green energy estimates. There is a risk of latency and single point failure due to the centralized job scheduler.

In order to strike a balance between two competing goals in data center networks—energy efficiency and fault tolerance—Shaukat et al., [16] suggested an Energy-Aware Fault-Tolerant (EAFT) Data Center Scheduler. Consider energy and data centers efficiency parameter of EAFT in relation to current energy efficiency strategies. On the other hand, the framework's inflexibility and lack of support for a customizable redundancy level make it unsuitable for use in highly dynamic cloud settings characterized by diverse rate of failure and resource heterogeneity.

A new scheduling approach based on carbon intensity was proposed by Yang et al. [17] for scheduling compute activities on clouds dynamically utilizing the drift-plus-penalty method of scheduling method in Lyapunov optimization. In order to assess the network's carbon footprint, a new model was suggested for the virtual queueing network, which records the network's computation and communication processes. A greedy technique was used to minimize the upper bound of drift-plus-penalty, and an efficient dynamic carbon intensity-based scheduling strategy was established. Though, it turns out that the minimization problem is an unbounded Knapsack problem that is NP-hard.

An eco-friendly framework for scheduling tasks in a federated cloud setting, Eco-friendly RL in Federated Cloud (ERLFC) was created by Wang et al., [18]. The Actor-Critic system sends tasks to certain data centers according to a number of criteria, including the present cloud data centers and the next task's total energy consumption, cooling technique, processing time, energy source type, emission ratio, and overall energy consumption. By utilizing the Heterogeneous Earliest Finish Time (HEFT) approach, tasks are directed to the most suitable server for processing.

The GAIA carbon-aware scheduler was developed by Hanafy et al. [19] to help users strike a compromise between the three trade-offs—carbon, performance, and cost—in cloud-based batch schedulers. GAIA uses flexible scheduling policies based on carbon and offers various cloud purchase models such as reserved, on-demand and spot instances. The scheduler uses the carbon intensity information and forecast to schedule batch jobs with carbon awareness. The implementation was tested using production workload traces and carbon intensity profiles, both available publicly, and from various geographical regions. However, this model has the high job completion time and higher operational cost.

To forecast low-carbon cloud computing loads, Zhang et al. [20] presented a CNN-BiLSTM model. This model combines convolutional neural networks with bidirectional long-term and short-term memory neural networks. In order to forecast the need for cloud computing, both a CNN and BiLSTM system together are used. An energy-based approach was taken to estimate future carbon emissions, with the help of cloud computing for real-time prediction. Additionally, in order to modify server carbon emission in relation to CPU utilization, a dynamic model was developed for predicting server carbon emissions. Predicted load fluctuates due to changes in the actual situation, leading to errors, even if the projected load value has nothing to do with the server's carbon emissions.

Neighbourhood Inspired Multiverse Scheduler (NIMS), developed by Tiwari et al., [21], is an innovative approach to task scheduling that attempts to enhance makespan and energy consumption, two measures that are inherently at odds with one another. As part of its solution improvements, NIMS improves upon the original Multiverse Optimiser (MVO) by implementing a new fitness-based neighbourhood search technique. If MVO's conventional update mechanism is underperforming, this improvement can assist improve solution quality even more. Despite this, the initial MVO's update mechanism for solutions simply makes random changes to current solutions without taking into account any problem-specific heuristic information on the optimisation parameters.

For power-aware and Service Level Agreement (SLA)-driven job scheduling in the cloud, Narsimhulu and Kumar [22] developed an RL-GA-LSTM-AE architecture. The LSTM is a system that autoencoder extraction method module is employed for the purpose of overload situation prediction and anomalous behaviors detection. Use of multi-objective evolutionary learning allowed the RL-GA to optimize job allocation. To be sure, AI-based systems may take longer to deploy real-time processing due to computing expense of training and inference.

Divya et al., [23] suggested the Quantum-Inspired Adaptive Meta-Heuristic-Machine Learning (QI-AMHML) technique for energy-efficient workload scheduling in multi-cloud systems. This model was created by combining the Whale Optimization Algorithm (WOA) with quantum inspired search dynamics to form a Reinforced Quantum WOA (RQWOA). Proactive placement decisions were made under varying load and fault conditions using a federated gradient boosting scheduler to improve resilience in multi-objective optimization, and solve energy and load balancing problems. However, existing scheduling methods often fail to handle highly dynamic workloads, leading to slow convergence or instability during unexpected anomalies.

A carbon-aware scheduling system was created by Danach et al., [24]. It incorporates adaptable rolling-horizon control, multi-objective optimization, and real-time estimates of carbon intensity. This model optimally distributed the computational workloads to leverage renewable energy availability, thereby reducing both operating costs and greenhouse gas emissions. The ensemble forecasting module uses gradient boosting regression and LSTM, and the framework uses a Mixed-Integer Linear Programming (MILP) framework that is solved using the ε-constraint technique. Workload arrivals and revised carbon estimates inform its scheduling decisions. The majority of current methods, however, rely on static or oversimplified models of carbon intensity rather than predictive analytics.

A cloud-based approach that can be used, as described by Beena et al., [25], integrates data on carbon intensity in real-

Paper ID: SR26827150359 DOI: https://dx.doi.org/10.21275/SR26827150359

time into the scheduling of energy-intensive tasks. Utilize the advantage of regional variations in carbon intensity in both historical and real-time analytics and use the AWS service to dynamically manage workloads with a high energy demand. It was designed to be easily extensible to existing cloud products such as AWS, Google Cloud and Azure. A scalable approach using Kubernetes-based scheduling and containerised workloads achieves intelligent resource management. However, it has the limited carbon reduction potential, complexity and overhead.

3. Comparative Analysis

Table 1: Green Cloud based Task Scheduling, Load Balancing and Fault-Tolerance

Ref no	Methods used	Merits	Demerits	Dataset	Performance Metrics
[11]	C-ACO	It reduces the makespan time while ensuring the load balancing	It higher throughput and low response time	Cloudsim	jobs vary from 500 to 1500 Makespan for (job size 80) = 577.31, standard deviation = 0.51. jobs vary from 1500 to 2000. Makespan for (job size 80)= 949.01, standard deviation = 0.53
[12]	Polynomial-time algorithm, DAG	It minimizes the makption	In exchange for higher energy consumption, it improves performance, allowing for quicker execution.	historical data	Heuristic rule: Energy consumption for = 4024.8wh, Energy saving = 0.00
[13]	DRL	Maximizing the utilization of renewable energy sources without time constraints	The infrastructure changes need model retraining	Real-world solar energy data	Energy consumption for jobs: 800 = 3254w, costs for 1000 jobs Cost = 2AU$
[14]	GWO	It improves convergence speed and reduce the local optima.	It has the high computational complexity.	Cloudsim	For inspiral_1000 Makespan = 4254, Cost = 17230, Energy Consumption = 55324w,
[15]	SG-LSTM, DMGC, GCDC	This effectively integrate green energy predictions with task scheduling	Potentially introducing delays and a singular point of failure, the centralized task scheduler	Wind speeds, solar irradiance, Google Cluster traces, and power costs	RMSE = 23.24, MAE = 11.03
[16]	EAFT	It reduces the energy consumption while maintain the fault tolerance	It uses the fixed one-level redundancy	Google open data set	Energy consumption by fail rate 20% = 168, Average throughput = 58,9674
[17]	Lyapunov optimization	Based on the current carbon intensity, it arranges the task dynamically.	Proved to be an NP-hard unbounded Knapsack problem, the minimization problem	ImageNet dataset	Carbon Emission = 53%
[18]	RL, Actor-Critic algorithm	It reduced the energy consumption and carbon emission	It has high computational complexity	Google Cluster dataset	No. of task = 472658 Energy consumption = 12709082.77 kw, carbon emission = 3825815
[19]	GAIA	It doubles the carbon saving per percentage increase in cost	this model has the high job completion time and higher operational cost.	Alibaba-PAI trace, Azure-VM trace, Los Alamos National Lab, Mustang-HPC trace	Carbon emission = 0.97 kgco2, Cost = 0.69, Waiting = 0.46
[20]	CNN and BiLSTM	The combined CNN-LSTM model has the good prediction efforts	Predicted load value does not match real server carbon emissions, causing cloud computing platform carbon emissions to not vary with load change, indicating an error.	Google open data set	MSE = 0.2222, MAE = 0.2387 and R^2 = 0.9494
[21]	NIMS, MVO	It improves the makespan, energy consumption, throughput	It undergoes from the random search updates	GoCJ, HPC2N, NASA	For the NIMS Makespan for GoCJ (task count 600) = 19.14, HPC2N (task count 2000) = 45.13, NASA = 2.65
[22]	RL-GA, LSTM	it improves the energy efficiency and resource utilization in dynamic cloud environment	It has the high computational complexity for training and inference	Google Cluster Traces and RUBiS	For: RL-GA Makespan = 99s, SLA Violation = 9.2%, Deadline Miss Rate = 9.3%, CPU utilization = 88.7%, Energy Consumption = 11.2kWh

[23]	QI-AMHML, WOA, RQWOA	Implements resilient and energy-efficient task scheduling in multi-cloud setups	Scheduling methods are slow to adapt in dynamic environments.	MultiCloudSynth-2025	QI-AMHML Energy consumption = 38.2%, Average service delay = 33.5%
[24]	LSTM	At the same time, it reduces operating expenses and emissions of greenhouse gases.	Current methods rely on static or oversimplified models of carbon intensity instead of dynamic forecasting.	Google Cloud workload traces and the UK National Grid	Emissions = 38,870 kg, cost = 375.8, MAPE = 6.0%
[25]	AWS, Google Cloud and Azure	In comparison to more conventional scheduling approaches, it vastly improves energy efficiency.	It has the complexity and overhead and dynamic carbon intensity data	CloudSim	-

Performance Analysis
This section evaluates the green cloud-based job scheduling, load balancing, and fault-tolerance in terms of Makespan and energy usage. The analysis highlights the capability of these frameworks to provide efficient task scheduling, balanced workload distribution, fault tolerance and reduced energy consumption in cloud environment.

Figure 4. shows that [21] has achieved better Makespan performance, among the other existing frameworks. The reduced makespan demonstrates the effectiveness of the framework in optimizing task scheduling, load balance and fault-tolerance across cloud resources and minimizing the execution delay. Furthermore, the [21] indicates the framework provides faster task completion and better resource utilization, making it highly suitable for the Green Cloud Computing environment.

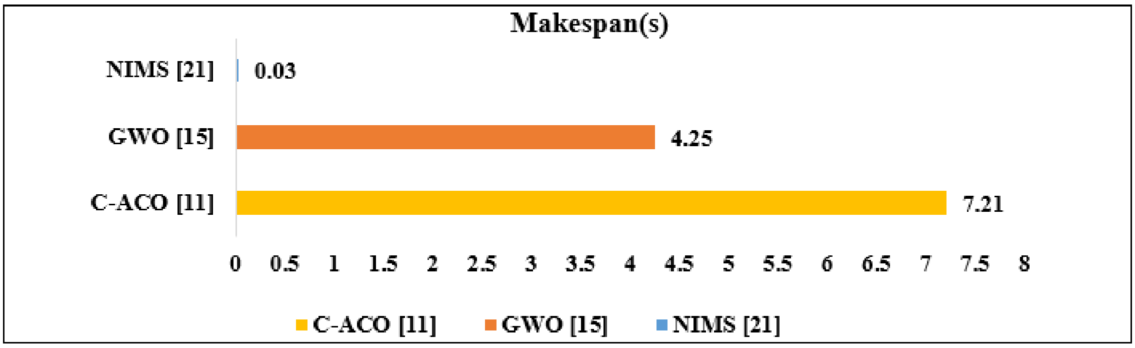

Figure 4: Comparative analysis of Makespan

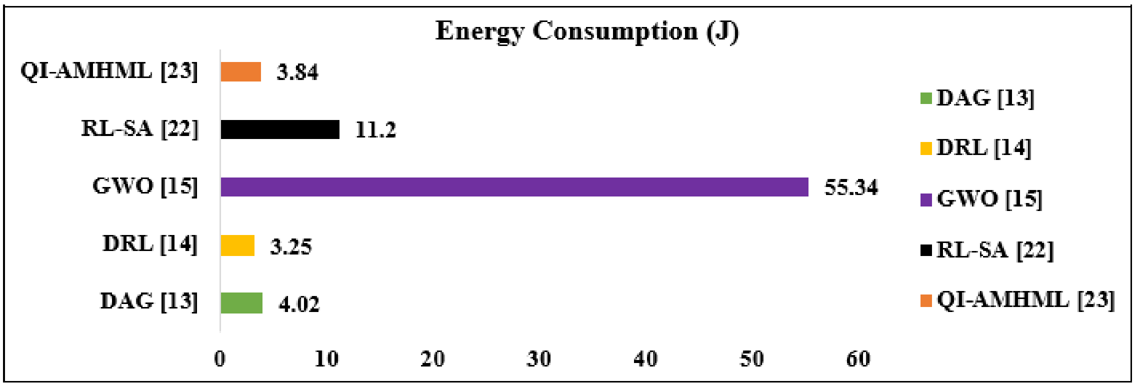

Figure 5: Evaluation of Energy Usage in comparison

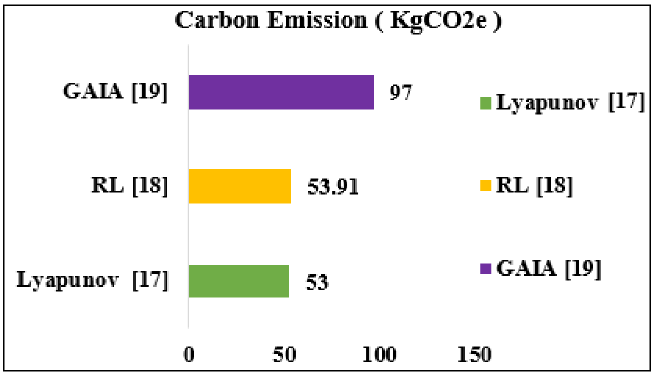

Figure 6: Comparison of Carbon Emission

Figure 5 displays the results of comparing the energy use with another existing model. The result show that the selected models have varying degrees of energy efficiency in the cloud environment. DRL [14] utilizes the least energy even among the compared approaches, while performing well in energy-aware task scheduling. In contrast, GWO [15] records the highest energy consumption of 55.34 kWh. The above analysis indicates that intelligent learning-based scheduling approaches have the potential to greatly decrease the energy consumption and enhance the sustainability of the Green Cloud Computing systems.

Paper ID: SR26827150359 DOI: https://dx.doi.org/10.21275/SR26827150359

Figure 6 shows the carbon emissions for the selected models of Green Cloud Computing. The results vary in the effectiveness of the various methods in lowering carbon emissions. Among the three, Lyapunov [17] gives the lowest carbon emitted value (53%), indicating that its task scheduling is more carbon-aware with more environmentally friendly. The carbon emission value reported in RL [18] is slightly higher (53.91%) than that of GAIA [19] which achieves the highest carbon emission reduction value (97%), thus showing the potential for carbon emissions reduction by implementing a carbon-aware scheduling strategy.

4. Discussion

The outcome of literature review shows that literatures [11]–[25] shows that there is a lack of efficient energy usage, increase in carbon emission, imbalanced load distribution and low reliability among various nodes in dynamic and heterogeneous cloud environments due to inefficiency of various scheduling algorithms in terms of carbon, load and fault tolerance. Future research involves the creation of holistic, adaptive scheduling systems that incorporate carbon-conscious decision making, incorporate intelligent load balancing mechanisms, and include robust fault-tolerant features, with the help of advanced AI and predictive analysis guiding the system in optimizing resource utilization, reducing carbon footprints, and ensuring traffic resilience and efficient cloud service delivery, even at the face of unpredictable workload fluctuations and infrastructure changes.

5. Conclusion

DL, RL and optimization are increasingly used in Green Cloud Computing to address issues of energy consumption, resource utilization, load balancing and fault tolerance. This research explored the advantages and disadvantages of the different Green Cloud job scheduling, load balancing, and fault-tolerant approaches and corresponding datasets and performance metrics. These approaches hold potential to enhance energy efficiency, reduce carbon footprints, enhance QoS, and ensure reliable cloud operations, with promising evidence and benefits. However, there remain important challenges that include dynamic workload management, making decisions on-the-fly in an environment of heterogeneous resources across the cloud, and addressing computational complexity. Future research is suggested to focus on the development of smart and adaptive framework DL, job scheduling, load balancing, and fault-tolerance mechanisms to enhance the sustainability, reliability and performance of the system.

References

[1] Kaur, T. (2019). Cloud computing: A study of the cloud computing services. international Journal for Research in Applied Science & Engineering Technology (IJRASET), 7(VI), 1933-1938.

[2] Amini, M., Sadat Safavi, N., Dashti Khavidak, S. M., & Abdollahzadegan, A. (2013). Types of cloud computing (public and private) that transform the organization more effectively. International Journal of Engineering Research & Technology (IJERT), 2(5), 1263-1269.

[3] Jing, S. Y., Ali, S., She, K., & Zhong, Y. (2013). State-of-the-art research study for green cloud computing. The Journal of Supercomputing, 65(1), 445-468.

[4] Goyal, V., & Garg, S. (2022). Green cloud computing for sustainable environment: A review. IOSR Journal of Environmental Science, Toxicology and Food Technology.

[5] Kumar, S., & Buyya, R. (2012). Green cloud computing and environmental sustainability. Harnessing green IT: principles and practices, 315-339.

[6] Singh, S., & Chana, I. (2016). A survey on resource scheduling in cloud computing: Issues and challenges. Journal of grid computing, 14(2), 217-264.

[7] Kumar, P., & Kumar, R. (2019). Issues and challenges of load balancing techniques in cloud computing: A survey. ACM computing surveys (CSUR), 51(6), 1-35.

[8] Kansal, N. J., & Chana, I. (2012). Cloud load balancing techniques: A step towards green computing. IJCSI International Journal of Computer Science Issues, 9(1), 238-246.

[9] Rehman, A. U., Aguiar, R. L., & Barraca, J. P. (2022). Fault-tolerance in the scope of cloud computing. Ieee Access, 10, 63422-63441.

[10] Bharany, S., Badotra, S., Sharma, S., Rani, S., Alazab, M., Jhaveri, R. H., & Gadekallu, T. R. (2022). Energy efficient fault tolerance techniques in green cloud computing: A systematic survey and taxonomy. Sustainable Energy Technologies and Assessments, 53, 102613.

[11] Ari, A. A. A., Damakoa, I., Titouna, C., Labraoui, N., & Gueroui, A. (2017, November). Efficient and scalable ACO-based task scheduling for green cloud computing environment. In 2017 IEEE International Conference on Smart Cloud (SmartCloud) (pp. 66-71). IEEE.

[12] Juarez, F., Ejarque, J., & Badia, R. M. (2018). Dynamic energy-aware scheduling for parallel task-based application in cloud computing. Future Generation Computer Systems, 78, 257-271.

[13] Zhao, J., Rodriguez, M. A., & Buyya, R. (2021, September). A Deep Reinforcement Learning Approach to Resource Management in Hybrid Clouds Harnessing Renewable Energy and Task Scheduling. In CLOUD (pp. 240-249).

[14] Mohammadzadeh, A., Masdari, M., Gharehchopogh, F. S., & Jafarian, A. (2021). Improved chaotic binary grey wolf optimization algorithm for workflow scheduling in green cloud computing. Evolutionary Intelligence, 14(4), 1997-2025.

[15] Bi-Objective, G. E. F. B. (2021). Scheduling of Tasks Across Distributed Clouds.

[16] Shaukat, M., Alasmary, W., Alanazi, E., Shuja, J., Madani, S. A., & Hsu, C. H. (2022). Balanced energy-aware and fault-tolerant data center scheduling. Sensors, 22(4), 1482.

[17] Yang, C. S., Huang-Fu, C. C., & Fu, I. K. (2022, December). Carbon-neutralized task scheduling for green computing networks. In GLOBECOM 2022-2022 IEEE Global Communications Conference (pp. 4824-4829). IEEE.

[18] Wang, Z., Chen, S., Bai, L., Gao, J., Tao, J., Bond, R. R., & Mulvenna, M. D. (2023). Reinforcement learning based task scheduling for environmentally sustainable

federated cloud computing. Journal of Cloud Computing, 12(1), 174.

[19] Hanafy, W. A., Liang, Q., Bashir, N., Souza, A., Irwin, D., & Shenoy, P. (2024, April). Going green for less green: Optimizing the cost of reducing cloud carbon emissions. In Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 3 (pp. 479-496).

[20] Zhang, H., Li, J., & Yang, H. (2024). Cloud computing load prediction method based on CNN-BiLSTM model under low-carbon background. Scientific Reports, 14(1), 18004.

[21] Tiwari, S., & Beena, B. M. (2024). A neighborhood inspired multiverse scheduler for energy and makespan optimized task scheduling for green cloud computing systems. IEEE Access, 12, 157272-157298.

[22] Narsimhulu, B., & Kumar, T. S. (2026). A hybrid RL–GA–LSTM–AE framework for energy-aware and SLA-driven task scheduling in cloud computing environments. Scientific Reports.

[23] Divya, N., Kiranbabu, M. N. V., & Babu, G. C. (2026). Quantum-Inspired Adaptive Meta-Heuristic–Machine Learning framework for resilient and energy-efficient task scheduling in multi-cloud ecosystems. Scientific Reports.

[24] Danach, K., Hamze, K., Harb, H., & Kanj, H. (2026). Carbon-Aware Scheduling in Cloud Computing Operations: A Multi-Objective Optimisation Approach. IET Smart Grid, 9(1), e70056.

[25] Beena, B. M., Ranga, P. C., Manideep, T. S. S., Saragadam, S., & Karthik, G. (2025). A green cloud-based framework for energy-efficient task scheduling using carbon intensity data for heterogeneous cloud servers. IEEE Access, 13, 73916-73938.