International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 2, Mar-Apr 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Carbon-Aware Load Balancing in Multi-Cloud Environments Using Real-Time Grid Emission Data

Omkar Bhoir*
*(Department of Information Technology, B K Birla College, Kalyan, Maharashtra, India
Email: bhoiromkar626@gmail.com
Under the Guidance of: Deepmala Maity & Vrunda Patil (Assistant Professors, Dept. of CS & IT))

****

Abstract:

Cloud computing's rapid growth imposes a growing environmental burden through data centre carbon emissions. The carbon footprint of any computational task is tied to the energy mix of its host grid — which varies significantly by geography and time. This paper proposes and evaluates a Carbon-Aware Load Balancing (CALB) framework for multi-cloud environments. CALB ingests real-time grid emission data (gCO_2eq/kWh) from the Electricity Maps API and dynamically routes workloads to the cloud region with the lowest current carbon intensity, subject to latency and cost constraints. Simulation across four major cloud regions using 2023 historical emission traces demonstrates 42% carbon reduction vs. round-robin scheduling, with a mean latency overhead under 18 ms and cost increase under 6%. Results confirm carbon-aware scheduling is both environmentally effective and operationally viable.

Index Terms—*carbon-aware computing, cloud load balancing, multi-cloud, green computing, grid carbon intensity, sustainable data centres, workload scheduling.*

I. INTRODUCTION

Cloud computing has become indispensable to modern digital infrastructure, underpinning everything from enterprise software to global social platforms. By 2030, AI-driven data centres alone are projected to account for approximately 8% of global electricity consumption . With electricity grids still heavily reliant on fossil fuels in many regions, this translates into a substantial and growing carbon footprint.

Existing cloud load balancing algorithms are designed almost exclusively around performance metrics: latency, throughput, availability, and cost. Carbon emissions remain an afterthought in workload placement decisions. Multi-cloud architectures — where enterprises simultaneously consume services from two or more cloud providers across geographically dispersed regions — create a structural opportunity to route workloads to cleaner grids.

This paper makes the following contributions:

(i) A Carbon-Aware Load Balancing (CALB) framework integrating real-time grid emission data (gCO_2eq/kWh) from the Electricity Maps API into multi-cloud workload routing.

(ii) A simulation-based evaluation across four cloud regions demonstrating up to 42% carbon reduction vs. round-robin baseline.

IV. EXPERIMENTAL EVALUATION

A. Setup

Experiments used a discrete-event simulation in Python 3.11 with SimPy, across four regions: us-east-1 (~400 gCO_2eq/kWh avg), europe-west1 (~150 gCO_2eq/kWh), ap-south-1 (~700 gCO_2eq/kWh), and ca-central-1 (~30 gCO_2eq/kWh). Historical 2023 hourly emission data (Electricity Maps API) drove the simulation. 10,000 synthetic tasks were generated with Poisson inter-arrivals (mean 30s) and log-normal durations (mean 5 min, $\sigma = 1.2$).

Three policies were evaluated: Round-Robin (RR), Lowest-Latency (LL), and the proposed CALB. All runs used the same NumPy random seed (42) for reproducibility.

B. Results

Table I summarises performance across all three policies. CALB achieved a 42.0% reduction in total carbon relative to RR (181.2 kg vs. 312.4 kg CO_2eq), primarily by routing flexible workloads to ca-central-1 (hydroelectric, lowest CI) and europe-west1 as secondary. The SLO violation rate for CALB (6.8%) fell between LL (4.1%) and RR (12.3%). The 18ms mean latency overhead reflects occasional routing to distant low-CI regions. Strict-SLO tasks always fell back to the nearest compliant region, ensuring zero strict-SLO violations from carbon routing. The 5.7% cost premium reflects regional compute pricing and inter-region transfer costs.

(iii) A discussion of design trade-offs including average vs. marginal emission signals, forecast uncertainty, and scalability.

II. RELATED WORK

A. Geographical Load Balancing

Liu et al. applied Lyapunov optimisation to jointly minimise electricity cost and carbon emissions across geo-distributed data centres while respecting SLA constraints. Their formulation demonstrated that spatial variability in grid carbon intensity could be substantially exploited for emission reduction. Zheng et al. showed that migrating workloads based on curtailment signals could simultaneously reduce carbon emissions and address renewable energy wastage.

B. Real-Time Carbon Signals in Orchestration

GreenCourier extended Kubernetes scheduling to incorporate real-time marginal emission scores per region via a custom plugin. CASPER jointly optimised distributed web-service provisioning and load-balancing under latency SLOs using a mixed-integer programme, treating carbon intensity as a first-class objective. Maji et al. modified VMware's Global Server Load Balancer to route traffic based on carbon emission factor (CEF, gCO_2eq/kWh).

C. Multi-Cloud Carbon Scheduling

A systematic review of 28 studies found only 11 combined spatial and temporal shifting, and only 2 developed a combined optimisation algorithm. MAIZX used an AI-driven agent approach, reporting up to 85% emission reduction. CarbonFlex applied case-based learning using historical carbon traces. Radovanovic et al. confirmed that greedy heuristics capture over 90% of theoretically achievable savings.

III. SYSTEM DESIGN

A. Problem Formulation

Let $R = \{r_1, r_2, ..., r_n\}$ be the set of n cloud regions, each with time-varying carbon intensity CI(ri, t) in gCO_2eq/kWh. Let $W = \{w_1, w_2, ..., w_m\}$ be incoming workloads, each characterised by compute duration d(wj), energy consumption E(wj), maximum latency L_max(wj), and cost ceiling C_max(wj). The CALB objective is:

$$\text{Minimise } \Sigma_j \; CI(A(wj), t) \times E(wj)$$
$$\text{Subject to: } latency(A(wj)) \leq L\_max(wj)$$
$$\text{and } cost(A(wj)) \leq C\_max(wj)$$

When all candidate regions satisfy constraints, CALB routes the workload to the minimum-CI region.

TABLE I
Performance Comparison of Scheduling Policies

Metric	RR	LL	CALB
Total Carbon	312 kg	289 kg	181 kg (−42%)
SLO Violation	12.3%	4.1%	6.8%
Mean Latency	148 ms	112 ms	130 ms
Cost Index	1.000	1.031	1.057

C. Seasonal and Diurnal Analysis

Carbon intensity varied significantly across seasons. us-east-1 exhibited ~22% higher average CI in summer vs. winter due to air conditioning load and peaker dispatch. europe-west1 showed strong diurnal variation (solar depression at midday). CALB dynamically adapted by updating its routing table every 15 minutes. During periods when all regions had elevated CI simultaneously (3.2% of simulation time), CALB defaulted to the global minimum-CI region, accepting higher cost overhead.

V. DISCUSSION

A. Effectiveness and Practicality

CALB's 42% carbon reduction is consistent with the upper range reported by spatial-shifting frameworks and substantially exceeds temporal-only approaches (typically 10–25%). The 5.7% cost overhead is the primary practical barrier to enterprise adoption. As carbon markets mature and cloud providers introduce carbon-linked pricing — as Microsoft and Google have begun exploring — this trade-off is expected to narrow.

B. Limitations

Limitations include: (i) synthetic rather than real production workloads; (ii) operational carbon only — embodied (hardware lifecycle) carbon excluded (iii) networking energy from inter-region transfer assumed negligible; and (iv) only four regions evaluated. A broader deployment would likely yield greater savings.

VI. CONCLUSION

This paper presented CALB, a provider-agnostic carbon-aware load balancing framework for multi-cloud environments. Evaluated via simulation over real 2023 emission traces, CALB achieved 42% carbon reduction vs. round-robin with an 18 ms latency overhead and 5.7% cost premium. Future work will extend CALB to combined spatio-temporal optimisation, integrate embodied carbon estimates, and deploy a reinforcement learning agent for adaptive weight tuning across production workloads.

Acknowledgment

The author thanks the Department of Information Technology, B K Birla College, Kalyan, and the project supervisor for guidance throughout this research.

Otherwise, a weighted penalty function across normalised carbon, latency, and cost scores determines assignment.

B. Architecture Overview

The CALB framework comprises four components: (a) Carbon Signal Collector — polls the Electricity Maps API at 15-minute intervals for gCO_2eq/kWh per region; (b) Workload Profiler — annotates tasks with duration, energy, latency class, and cost tier; (c) Carbon-Aware Scheduler — core routing engine implementing the objective function; and (d) Telemetry Module — logs per-task routing decisions for post-hoc analysis.

C. Carbon Signal Selection

Average carbon intensity (rather than marginal) was selected as the primary signal, as provided natively by the Electricity Maps API. While Sukprasert et al. [11] show marginal signals theoretically outperform average (~3% savings lost per 14% forecast error), average signals offer broader coverage and higher reliability across heterogeneous multi-cloud markets — appropriate for a production-oriented prototype.

References

[1] P. Wadhwani, "AI Data Centers and Global Electricity Demand," Energy Research Reports, 2025.

[2] Z. Liu et al., "Carbon-Aware Load Balancing for Geo-Distributed Cloud Services," in Proc. IEEE MASCOTS, San Francisco, 2013.

[3] J. Zheng, A. A. Chien, and S. Suh, "Mitigating Curtailment and Carbon Emissions through Load Migration Between Data Centers," Joule, vol. 4, no. 10, 2020.

[4] M. Chadha et al., "GreenCourier: Carbon-Aware Scheduling for Kubernetes," in Proc. ACM SoCC, 2023.

[5] A. Souza et al., "CASPER: Carbon-Aware Spatial Provisioning and Routing," in Proc. ACM SIGCOMM, 2024.

[6] A. Maji et al., "Bringing Carbon Awareness to Multi-Cloud Application Delivery," in Proc. ACM IMC, 2023.

[7] M. Georgiadis et al., "Carbon-Aware Spatio-Temporal Workload Shifting: A Review," Sustainability, vol. 17, no. 14, 2025.

[8] J. Ruilova Alfaro, "MAIZX: A Carbon-Aware Framework for Cloud Emissions," arXiv:2506.19972, 2025.

[9] W. Hanafy et al., "CarbonFlex: Case-Based Learning for Carbon-Aware Scheduling," in Proc. HotCarbon, 2025.

[10] A. Radovanovic et al., "Carbon-Aware Computing for Datacenters," IEEE Trans. Power Syst., vol. 38, no. 2, pp. 1270–1280, 2022.

[11] T. Sukprasert et al., "On the Implications of Choosing Average vs. Marginal Carbon Signals," in Proc. HotCarbon, 2023.

[12] R. Hischier et al., "Grey Energy and Environmental Impacts of ICT Hardware," in ICT Innovations for Sustainability, Springer, 2015.

