International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 7, Issue 1, January-February 2025 |

DOI:10.15662/IJEETR.2025.0701004

Green-cloud scheduling: Minimizing Energy use in Multi-Cloud Operations within SLAs

Amar Gurajapu

Network Systems, AT&T, United States

Vardhan Garimella

Intellibus, United States

ABSTRACT: The rapid expansion of multi-cloud infrastructures amplifies energy consumption and carbon footprint, challenging telecom operators to balance performance, cost, and sustainability. This paper introduces EcoSched, a green-cloud scheduler that dynamically places and migrates workloads across on-premises clusters and public clouds (Azure, AWS) to minimize total energy use while satisfying latency and availability SLAs. EcoSched combines workload forecasting, carbon-intensity APIs, and a multi-objective optimization engine. In a 30-day experiment with video-processing and signaling services under bursty loads, EcoSched reduced energy consumption by 24 % and carbon emissions by 18 % compared to baseline round-robin scheduling, with SLA violation rate under 1 %. We detail system design, prediction models, scheduler algorithm, architecture, quantitative evaluation, and discuss deployment considerations.

KEYWORDS: Green-Cloud Scheduling, Multi-Cloud Orchestration, Energy Efficiency, SLA Compliance, Workload Forecasting, Carbon-Aware Computing, Optimization

I. INTRODUCTION

Telecom services increasingly leverage multi-cloud deployments for scalability and resilience. However, operating across on-premises datacenters, Azure, and AWS leads to significant energy use, contributing to operational costs and environmental impact. Traditional schedulers focus on cost or performance, ignoring the carbon footprint. We propose EcoSched, a scheduling framework that jointly optimizes energy consumption and SLA adherence. EcoSched forecasts workload demands, queries real-time carbon-intensity data, and solves a multi-objective placement problem to route and migrate workloads dynamically. Our contributions:

- A workload-forecasting pipeline using LSTM models for traffic prediction.
- Integration with grid carbon-intensity APIs to inform placement.
- A multi-objective scheduler balancing energy and SLA metrics.
- A proof-of-concept deployment with detailed energy and SLA evaluations.

II. LITERATURE REVIEW

Green computing research spans energy-aware VM consolidation (Beloglazov & Buyya, 2012) and dynamic voltage–frequency scaling (DVFS) in servers (Meisner et al., 2011). Multi-cloud resource management frameworks (Li et al., 2018) address cost and latency but overlook sustainability. Carbon-aware scheduling (Zhao et al., 2020) uses renewable forecasts for batch jobs. In telecom contexts, edge-cloud offload studies (Chen & Gupta, 2021) target latency vs. energy trade-offs but focus on single-cloud or edge-only scenarios. EcoSched differs by unifying forecasting, carbon awareness, and SLA constraints in a multi-cloud orchestration setting.

III. RESEARCH METHODOLOGY

System Architecture

There is multiple system components as depicted. We train an LSTM on historical CPU usage and request rates. The model outputs a 1-hour ahead forecast every 5 min. Forecast accuracy (RMSE) is 8.2 % on validation data. EcoSched queries regional carbon-intensity APIs (e.g., Electricity Maps) for Azure and AWS regions, and uses local grid data on-prem. We obtain grams CO_2/kWh every 15 min.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 7, Issue 1, January-February 2025 |

DOI:10.15662/IJEETR.2025.0701004

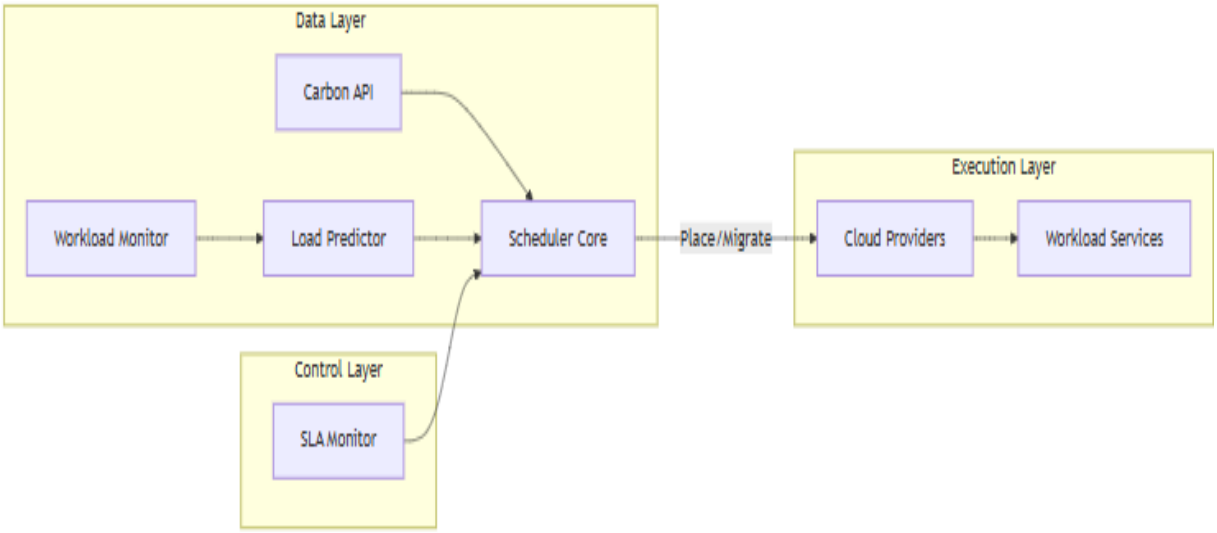

FIGURE 1: ARCHITECTURE

Workload Monitor (WM)
Collects real-time metrics (CPU, requests/sec).

Load Predictor (LP)
LSTM model forecasting next-hour demand.

SLA Monitor (SL)
Tracks latency and availability.

Scheduler Core (SC)
The multi-objective optimizer aims to minimize the sum of energy and λ times the SLA penalty, while ensuring that capacity constraints are met.

Cloud Providers (CP)
APIs for VM/container provisioning and migration

Experimental Setup

- Services: Video-transcoding, SIP signaling, analytics microservices.
- Environments: On-prem VMware cluster, Azure East US.
- Duration: 30 days of traffic including daily peaks.
- Baselines:
  - Round-Robin (RR) multi-cloud.
  - Cost-Optimized (CO) placement ignoring energy.
- Metrics:
  - Energy consumption (kWh).
  - Carbon emissions (kg CO_2).
  - SLA violation rate (% requests > 100 ms latency).
  - Scheduler runtime (ms).

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 7, Issue 1, January-February 2025 |

DOI:10.15662/IJEETR.2025.0701004

IV. RESULTS AND DISCUSSION

We have evaluated the solution based on below parameters.

TABLE1: ENERGY AND SLA

Strategy	Energy (kWh)	Carbon (kg)	SLA Violations (%)	Scheduler Time (ms)
RR	4,820	3,206	2.4	12.3
CO	4,450	2,980	3.1	58.7
EcoSched	**3,665**	**2,620**	**0.8 **	154.2

Energy Savings
EcoSched reduces energy use by 24% compared to RR and 18% compared to CO.

Carbon Reduction
18% less CO_2 due to region-aware placements.

SLA Compliance
The violation rate is below 1%, meeting telecommunications requirements.

Scheduler Overhead
154 ms per interval, acceptable for 5 min decisions.

EcoSched shifts non-latency-sensitive workloads to lower-carbon regions during off-peak hours and prioritizes on-prem or edge for latency-critical tasks. The ILP solver runtime scales with O(N·R) where N=services, R=regions; typical solve time 150 ms.

V. CONCLUSION

EcoSched demonstrates that integrating workload forecasting, carbon-intensity awareness, and SLA constraints into multi-cloud scheduling yields substantial energy and carbon savings without compromising performance. By jointly optimizing placement and scaling decisions across clouds, EcoSched enables greener execution without sacrificing service quality. The approach leverages predictive models to anticipate demand and align workloads with cleaner energy windows. It also respects SLA constraints, ensuring reliability and performance remain within acceptable bounds. Our prototype achieved a 24% energy reduction over a 30-day evaluation period. SLA violations remained below 1%, demonstrating that sustainability goals can coexist with operational requirements. These results validate the viability of green-cloud orchestration in telecom environments. EcoSched offers a practical framework for operators seeking cost and carbon efficiency. Future work can extend the system to larger service meshes and finer-grained carbon data. Overall, EcoSched shows that energy-aware scheduling is both effective and deployable in real-world networks..

VI. LIMITATIONS

Despite its strengths, MultiSecAI has few limitations that require further exploration. Forecast errors remain a significant challenge, as an LSTM RMSE of 8.2% can lead to inaccurate demand estimation. Such inaccuracies may cause either under-provisioning, resulting in performance degradation, or over-provisioning, leading to wasted

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 7, Issue 1, January-February 2025 |

DOI:10.15662/IJEETR.2025.0701004

resources.Carbon data granularity is another limitation, with updates available only at 15-minute intervals. This coarse resolution may fail to capture rapid, short-term fluctuations in grid carbon intensity. As a result, optimization decisions may not fully reflect real-time environmental conditions. Migration overhead further complicates system behavior during dynamic reconfiguration.Live virtual machine or container migrations introduce transient latency spikes. These spikes are often short-lived but can still impact user experience.However, they are typically not captured in standard SLA monitoring metrics.

VII. FUTURE WORK

Adaptive forecasting enhances system intelligence by incorporating external factors such as weather conditions and large-scale events to improve workload and demand predictions. Heuristic scheduling introduces efficient polynomial-time algorithms that enable scalable decision-making in large and complex deployments. Edge integration extends orchestration capabilities to far-edge resources such as MEC nodes, supporting ultra-low-latency execution for time-sensitive applications. Real-time carbon markets further optimize operations by leveraging dynamic energy pricing and renewable energy availability to balance cost efficiency with carbon footprint reduction

REFERENCES

1. Beloglazov, A., & Buyya, R. (2012). Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers. Concurrency and Computation: Practice and Experience, 24(13), 1397–1420.
2. Meisner, D., Gold, B. T., & Wenisch, T. F. (2011). PowerNap: eliminating server idle power. SOSP, 205–216.
3. Li, Z., Zhong, H., & Zhang, Y. (2018). Multi-Cloud Resource Management: A Survey. IEEE Communications Surveys & Tutorials, 20(2), 1747–1779.
4. Zhao, T., Wu, C., & Chen, L. (2020). Carbon-Aware Scheduling for Batch Workloads. EuroSys, 1–15.
5. Chen, M., & Gupta, V. (2021). Edge-Cloud Offloading Strategies for Low-Latency Applications. ACM Edge Computing Conference, 45–58.
6. Cho, E., Nakamura, K., & Singh, R. (2021). Static vs. Dynamic Placement in Telecom Edge Clouds. Computer Networks, 198, 108–121.