American Journal of Applied Science and Technology

VOLUME Vol. 06 Issue 03 2026
PAGE NO. 125-129

Carbon-Aware Multi-Objective Scheduling and Optimization in Geo-Distributed Cloud and Machine Learning Systems: A Sustainable Computing Paradigm

Maria Chokwuk

Department of Computer Science, University of Edinburgh, United Kingdom

Received: 01 January 2026; **Accepted:** 16 February 2026; **Published:** 31 March 2026

Abstract: The rapid proliferation of artificial intelligence (AI), cloud computing, and large-scale machine learning (ML) systems has significantly increased global computational demand, raising urgent concerns about energy consumption and carbon emissions. While recent advancements in model architectures and hardware acceleration have improved computational efficiency, the environmental footprint of these technologies remains substantial. This study investigates the integration of carbon-aware strategies into scheduling and optimization frameworks for geographically distributed cloud environments. Drawing upon contemporary research in carbon measurement, federated learning, workload migration, and evolutionary optimization, this work proposes a comprehensive theoretical model for multi-objective scheduling that simultaneously optimizes performance, cost, and carbon emissions. The study critically examines existing approaches such as carbon-aware federated learning, virtual machine placement, and checkpointing mechanisms, and contextualizes them within broader scheduling paradigms including heuristic, metaheuristic, and evolutionary algorithms. Furthermore, the research highlights the role of real-time carbon monitoring systems and geo-distributed infrastructure in enabling dynamic decision-making. The findings suggest that integrating carbon-awareness into scheduling algorithms can significantly reduce emissions without compromising computational performance. However, challenges such as data heterogeneity, system latency, and trade-off management persist. This paper contributes to the growing body of sustainable computing literature by offering an in-depth theoretical synthesis and proposing future research directions for scalable, environmentally responsible computing systems.

Keywords: Carbon-aware computing, multi-objective scheduling, cloud computing, machine learning sustainability, geo-distributed systems, evolutionary optimization, energy efficiency

INTRODUCTION

The exponential growth of computational technologies, particularly in the domains of artificial intelligence and cloud computing, has transformed modern society. From autonomous systems to large-scale data analytics, the reliance on high-performance computing infrastructure has become indispensable. However, this technological advancement has come at a significant environmental cost. The energy consumption associated with data centers, training large machine learning models, and maintaining cloud infrastructure has raised serious concerns about sustainability and carbon emissions. Studies have shown that the environmental footprint of computing is not only substantial but also complex, involving multiple layers of hardware, software, and geographical factors (Gupta et al., 2021).

The development of large-scale models such as those described in the GPT-4 technical report illustrates the immense computational resources required for modern AI systems (Achiam et al., 2023). These models often require extensive training on distributed hardware, consuming vast amounts of electricity. While some research suggests that the carbon footprint of machine learning training may eventually plateau due to efficiency improvements (Patterson et al., 2022), the current trajectory indicates a growing environmental impact. This is further compounded by the increasing demand for real-time inference and continuous model updates.

A critical aspect of addressing this challenge lies in understanding and measuring the carbon intensity of computational processes. Recent work has focused on quantifying the emissions associated with cloud-based AI workloads, highlighting the variability in carbon intensity across different regions and time periods (Dodge et al., 2022). Additionally, real-time monitoring systems such as CarbonMonitor-Power provide granular insights into global power generation and emissions, enabling more informed decision-making (Zhu et al., 2023).

Despite these advancements, there remains a significant gap in integrating carbon-awareness into the core scheduling and optimization mechanisms of computing systems. Traditional scheduling algorithms in cloud and grid environments have primarily focused on performance metrics such as execution time, resource utilization, and cost (Mishra et al., 2014; Tsai & Rodrigues, 2014). While these factors are crucial, they do not account for the environmental impact of computational decisions. This limitation has led to the emergence of carbon-aware scheduling as a promising research direction.

Carbon-aware scheduling involves optimizing the allocation and execution of computational tasks based on the carbon intensity of energy sources. This approach leverages geographical and temporal variations in energy production to minimize emissions. For instance, workloads can be shifted to regions with lower carbon intensity or scheduled during periods of renewable energy availability. Recent studies have explored this concept in various contexts, including federated learning (Bian et al., 2024), workload migration (Park et al., 2024), and virtual machine placement (Khodayarseresht et al., 2023).

However, implementing carbon-aware strategies introduces new complexities. Scheduling decisions must now balance multiple objectives, including performance, cost, reliability, and environmental impact. This multi-objective nature necessitates the use of advanced optimization techniques, such as evolutionary algorithms and metaheuristics. Algorithms like NSGA-II and MOPSO have been widely used for multi-objective optimization, offering robust solutions for complex scheduling problems (Deb et al., 2002; Coello Coello & Lechuga, 2002).

This study aims to provide a comprehensive theoretical framework for carbon-aware multi-objective scheduling in geo-distributed cloud environments. By synthesizing insights from existing literature and exploring the interplay between different optimization strategies, this research seeks to address the limitations of current approaches and propose scalable solutions for sustainable computing.

METHODOLOGY

The methodological framework of this study is grounded in a theoretical synthesis of existing research on carbon-aware computing, scheduling algorithms, and multi-objective optimization. Rather than relying on empirical experimentation, this work adopts a conceptual modeling approach to integrate diverse strands of literature into a unified framework.

The first step involves analyzing the characteristics of geo-distributed cloud environments. These systems consist of multiple data centers located across different geographical regions, each with varying energy sources, carbon intensities, and resource capacities. The heterogeneity of these environments presents both challenges and opportunities for carbon-aware scheduling. By leveraging spatial and temporal variations in energy availability, it is possible to optimize workload distribution in a way that minimizes emissions.

The second component focuses on carbon measurement and monitoring. Accurate and real-time data on carbon intensity is essential for informed decision-making. Systems such as CarbonMonitor-Power provide high-resolution data on power generation and emissions, enabling dynamic scheduling strategies (Zhu et al., 2023). This data is integrated into the scheduling framework to guide task allocation decisions.

The third aspect involves the formulation of the scheduling problem as a multi-objective optimization task. The objectives include minimizing execution time, reducing energy consumption, lowering carbon emissions, and maintaining system reliability. These objectives are often conflicting, requiring trade-offs. For example, shifting workloads to a low-carbon region may increase latency or cost. To address this, the study explores the use of evolutionary algorithms such as NSGA-II, which are capable of identifying Pareto-optimal solutions (Deb et al., 2002).

In addition to evolutionary algorithms, the methodology incorporates heuristic and metaheuristic approaches. Traditional scheduling heuristics, such as greedy algorithms and fuzzy prediction models, provide baseline solutions that can be enhanced with carbon-awareness (Zhou & Hu, 2014; Kong et al., 2011). Metaheuristic techniques, including particle swarm optimization and ant colony optimization, offer flexibility in exploring large solution spaces (Coello Coello & Lechuga, 2002; Lopez-Ibanez & Stutzle, 2012).

The framework also integrates advanced techniques such as checkpointing and workload migration. Efficient checkpointing systems enable the interruption and resumption of tasks, facilitating dynamic scheduling based on changing carbon conditions (Xu et al., 2024). Similarly, workload migration strategies allow tasks to be transferred between data centers to optimize carbon efficiency (Park et al., 2024).

Finally, the methodology considers the role of federated learning in distributed environments. Carbon-aware federated learning approaches aim to optimize the training process by selecting participants based on their carbon footprint (Bian et al., 2024). This adds another layer of complexity to the scheduling problem, as it involves coordination across multiple devices and networks.

RESULTS

The theoretical analysis reveals several key findings regarding the effectiveness of carbon-aware scheduling in geo-distributed systems. First, the integration of real-time carbon intensity data significantly enhances the ability to make environmentally optimized decisions. By aligning computational tasks with periods of low carbon intensity, it is possible to achieve substantial reductions in emissions without compromising performance.

Second, the use of multi-objective optimization techniques enables a balanced approach to scheduling. Evolutionary algorithms such as NSGA-II demonstrate strong potential in identifying solutions that simultaneously optimize multiple criteria. These algorithms are particularly effective in handling the trade-offs inherent in carbon-aware scheduling.

Third, the incorporation of workload migration and checkpointing mechanisms enhances system flexibility. These techniques allow for dynamic adaptation to changing conditions, enabling more efficient use of resources. For example, tasks can be paused and resumed in regions with lower carbon intensity, reducing overall emissions.

Fourth, carbon-aware federated learning emerges as a promising approach for distributed AI systems. By selecting participants based on their carbon footprint, it is possible to reduce the environmental impact of training processes. However, this approach also introduces challenges related to data heterogeneity and communication overhead.

Fifth, the analysis highlights the importance of integrating carbon-awareness into existing scheduling frameworks. Rather than developing entirely new systems, it is often more practical to enhance existing algorithms with carbon-related parameters. This approach leverages the strengths of established methods while addressing environmental concerns.

DISCUSSION

The findings of this study underscore the importance of adopting a holistic approach to sustainable computing. Carbon-aware scheduling represents a paradigm shift in the way computational resources are managed, emphasizing environmental impact alongside traditional performance metrics. However, the implementation of such strategies is not without challenges.

One of the primary limitations is the availability and accuracy of carbon intensity data. While systems like CarbonMonitor-Power provide valuable insights, there are still gaps in coverage and resolution. Inaccurate data can lead to suboptimal decisions, undermining the effectiveness of carbon-aware scheduling.

Another challenge lies in managing trade-offs between competing objectives. For instance, minimizing carbon emissions may conflict with performance or cost requirements. This necessitates the development of sophisticated optimization techniques capable of balancing multiple criteria. While evolutionary algorithms offer promising solutions, they can be computationally intensive and may not scale well for large systems.

The heterogeneity of geo-distributed environments also presents difficulties. Differences in hardware, network conditions, and energy sources complicate the scheduling process. Addressing these challenges requires a deep understanding of system dynamics and the development of adaptive algorithms.

Furthermore, the integration of carbon-aware strategies into existing infrastructure requires significant changes in system design and operation. This includes modifications to scheduling policies, data management systems, and communication protocols. The transition to carbon-aware computing is therefore not only a technical challenge but also an organizational one.

Despite these challenges, the potential benefits of carbon-aware scheduling are substantial. By reducing emissions and improving energy efficiency, these strategies contribute to the broader goal of sustainable development. Moreover, they align with increasing regulatory and societal pressures to reduce environmental impact.

Future research should focus on developing scalable and efficient optimization algorithms, improving carbon measurement systems, and exploring new applications of carbon-aware computing. In particular, the integration of machine learning techniques into scheduling algorithms offers exciting possibilities for adaptive and intelligent decision-making.

CONCLUSION

This study provides a comprehensive theoretical exploration of carbon-aware multi-objective scheduling in geo-distributed cloud and machine learning systems. By synthesizing insights from a wide range of literature, it highlights the importance of integrating environmental considerations into computational decision-making processes.

The analysis demonstrates that carbon-aware scheduling can significantly reduce emissions while maintaining performance and cost efficiency. However, achieving this requires advanced optimization techniques, accurate data, and adaptive system design. While challenges remain, the continued development of carbon-aware strategies represents a critical step toward sustainable computing.

As the demand for computational resources continues to grow, the need for environmentally responsible solutions becomes increasingly urgent. This research contributes to the ongoing effort to address this challenge, providing a foundation for future innovations in carbon-aware computing.

REFERENCES

1. Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al. GPT-4 technical report. arXiv preprint arXiv:2303.08774 (2023).

2. Gupta, U., Kim, Y.G., Lee, S., Tse, J., Lee, H.-H.S., Wei, G.-Y., Brooks, D., Wu, C.-J. Chasing carbon: the elusive environmental footprint of computing. IEEE International Symposium on High-Performance Computer Architecture (2021), pp. 854–867.

3. Patterson, D., Gonzalez, J., Hölzle, U., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., So, D.R., Texier, M., Dean, J. The carbon footprint of machine learning training will plateau, then shrink. Computer, 55(7) (2022), pp. 18–28.

4. Dodge, J., Prewitt, T., Tachet des Combes, R., Odmark, E., Schwartz, R., Strubell, E., Luccioni, A.S., Smith, N.A., DeCario, N., Buchanan, W. Measuring the carbon intensity of AI in cloud instances. Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (2022), pp. 1877–1894.

5. Zhu, B., Deng, Z., Song, X., Zhao, W., Huo, D., Sun, T., Ke, P., Cui, D., Lu, C., Zhong, H., et al. Carbonmonitor-power near-real-time monitoring of global power generation on hourly to daily scales. Scientific Data, 10(1) (2023), p. 217.

6. Bian, J., Wang, L., Ren, S., Xu, J. CAFE: carbon-aware federated learning in geographically distributed data centers. Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems (2024), pp. 347–360.

7. Xu, W., Huang, X., Meng, S., Zhang, W., Guo, L., Sato, K. An efficient checkpointing system for large machine learning model training. SC24-W: Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis (2024), pp. 896–900.

8. Park, J., Kim, D., Kim, J., Han, J., Chun, S. Carbon-aware and fault-tolerant migration of deep learning workloads in the geo-distributed cloud. IEEE International Conference on Cloud Computing (2024), pp. 494–501.

9. Khodayarseresht, E., Shameli-Sendi, A., Fournier, Q., Dagenais, M. Energy and carbon-aware initial VM placement in geographically distributed cloud

data centers. Sustainable Computing: Informatics and Systems, 39 (2023), Article 100888.

10. Mishra, M.K., Patel, Y.S., Rout, Y., Mund, G.B. A survey on scheduling heuristics in grid computing environment. International Journal of Modern Education and Computer Science, pp. 57–83 (2014).

11. Tsai, C., Rodrigues, J. Meta heuristic scheduling for cloud: A survey. IEEE Systems Journal, 8(1), 279–291 (2014).

12. Zhou, Z., Hu, Z. Task scheduling algorithm based on greedy strategy in cloud computing. Open Cybernetics and Systemics Journal, 8, 111–114 (2014).

13. Kong, X., Lin, C., Jiang, Y., Yan, W., Chu, X. Efficient dynamic task scheduling in virtualized data centers with fuzzy prediction. Journal of Network and Computer Applications, 34(4), 1068–1077 (2011).

14. Sun, W., et al. A game theoretic resource allocation model based on extended second price sealed auction in grid computing. Journal of Computers, 7(1), 65–75 (2012).

15. Grover, R., Chabbra, A. Bio-inspired optimization techniques for job scheduling in grid computing. IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (2016), pp. 1902–1906.

16. Bagchi, T.P. The nondominated sorting genetic algorithm: NSGA. In Multiobjective Scheduling by Genetic Algorithms. Springer (1999).

17. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 6(2), 182–197 (2002).

18. Coello Coello, C.A., Lechuga, M.S. MOPSO: a proposal for multiple objective particle swarm optimization. Proceedings of the 2002 Congress on Evolutionary Computation (2002), pp. 1051–1056.

19. Li, H., Landa-Silva, D. An adaptive evolutionary multi-objective approach based on simulated annealing. Evolutionary Computation, pp. 561–595 (2011).

20. Lopez-Ibanez, M., Stutzle, T. The automatic design of multiobjective ant colony optimization algorithms. IEEE Transactions on Evolutionary Computation, 16(6), 861–875 (2012).

21. Zhou, A., Qu, B.-Y., Li, H., Zhao, S.-Z. Multiobjective evolutionary algorithms: A survey of the state of the art. Swarm and Evolutionary Computation, 1(1), 32–49 (2011).

22. Yang, S., et al. A grid-based evolutionary algorithm for many-objective optimization. IEEE Transactions on Evolutionary Computation, 17(5), 721–736 (2013).

23. S. Bhat, S. R. Sirikonda, V. Katoch and R. Jain, "Carbon-Kube: A Kubernetes-Native Framework for Multi-Objective Carbon-Aware Scheduling of Big Data Pipelines," 2026 9th International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech), Kolkata, India, 2026, pp. 1-6, doi: 10.1109/IEMENTech202669403.2026.11434192.