Maximizing Return On Investment for Sustainable Operations through Smart Workload Migration

Georgia Christofidi*
Telefónica Research
IMDEA Software Institute
Universidad Politécnica de Madrid

Francisco Álvarez Terribas
Jesus Alberto Omaña Iglesias
Nicolas Kourtellis†
Telefónica Research

Thaleia Dimitra Doudali
IMDEA Software Institute

***Abstract*—The push for sustainable computing has led to an increased adoption of carbon-aware workload migration, where applications are shifted to time periods and regions with lower carbon intensity. Recently proposed solutions rely on the existence of infrastructures across multiple countries, assuming access to greener locations. However, this does not reflect the reality of national companies that only have on-premise IT infrastructure. For such companies to become more sustainable, they must inevitably increase their monetary expenses and expand their native or cloud infrastructure to other countries. Our analysis quantifies this trade-off between sustainability and cost: while migrating workloads to greener regions reduces carbon emissions by an order of magnitude, it also doubles deployment costs. Yet, we uncover a key insight: the application type and characteristics, such as latency and storage footprint, significantly impact carbon emissions and migration costs. This highlights the need for novel solutions that maximize returns on investments.**

I. Overview

The growing push for sustainable computing has driven the design of new resource management and workload execution techniques [1]–[6], that shift workload execution in times and locations where the carbon intensity of electricity generation [7] allows for reduced overall carbon emissions. More specifically, spatial shifting moves tasks to regions with greener energy sources [3], such as Sweden or Nepal [7], while temporal shifting delays non-urgent workloads [1], such as batch jobs [4], until there is access to greener, renewable energy sources, such as solar energy. The above carbon-aware techniques are increasingly integrated with traditional resource management techniques, prioritizing the reduction of emissions at the potential expense of performance implications or increased operational costs [2].

However, the above carbon optimization strategies **are viable only for multinational** cloud providers and large companies with data centers that span multiple countries. National companies have infrastructure within few countries, typically one, whose carbon intensity and energy sources may not be very 'green' [7]. Such companies can reduce their carbon emissions with spatial shifting by expanding their native or cloud infrastructure to greener countries, which incurs substantial monetary costs. Similarly, temporal shifting is not possible for many real-world, latency-sensitive applications that these companies host. Lastly, compliance with data privacy regulations, such as GDPR, often prohibits offloading data outside of regions, such as the European Union.

The above limitations present a trade-off: while carbon-aware workload execution can reduce emissions, it is applicable only for a limited set of applications (e.g., long-lasting batch jobs [2]) and incurs additional cost for expanding infrastructure to greener countries. We then ask; **how can a national company operate with reduced carbon emissions, in return for minimal cost and uninterrupted user service and satisfaction?** In this work, we preliminarily quantify the carbon-cost trade-off that national companies face, when they rent or buy remote resources, in order to make application deployment greener. We consider a realistic scenario of microservice applications with user traffic that is first processed in a local data center before being redirected to a remote, lower-carbon region. Next, we present the details of our experimental characterization.

Motivational experiment. To analyze the trade-off between carbon efficiency and cost, we consider a company that deploys its entire cloud-edge infrastructure in Spain, a country with a moderate carbon intensity of 206 grams of CO2 equivalent per kilowatt-hour (gC02eq/kWh) [7]. Recent works [1], [4], [5] suggest offloading workloads to 'greener' locations, such as Sweden [7], where carbon intensity is only 20 gC02eq/kWh. Our study aims to capture the difference in the total carbon footprint and the cost ($) of deploying a workload in these 2 countries, as well as the additional cost ($) needed for moving the workload from Spain to Sweden.

Performance of microservice applications. We evaluate two microservice-based applications from DeathStarBench [8], [9]: a social network application with 24 microservices and a media streaming application with 32 microservices. Each application runs on a separate node, with Kubernetes as a cluster manager and Prometheus for monitoring. We deploy a workload that sends 1,000 requests to each application at time steps that follow a Poisson distribution, emulating multiple concurrent users for a duration of 10 minutes.Table I summarizes the performance results of the 2 different microservice applications. We observe that the media streaming application has $2.89\times$ higher average latency than the social network

*Work done while Georgia Christofidi was at Telefonica Research. Currently she is a PhD student at the IMDEA Software Institute and Universidad Politécnica de Madrid.

† Nicolas Kourtellis is currently at Keysight.

Application	AVG Lat	P95/P99 Lat	Storage	Memory
SocialNet	9.49 ms	42.12/89.16 ms	0.85 GB	5.15 GB
MediaStream	26.08 ms	70.65/127.52 ms	0.75 GB	3.71 GB

TABLE I: Application performance regardless of location.

App(Location)	Carbon (mgCO2eq)	Local($/hr) + Move ($)
SocialNet(**ES**)	72.72	0.0912 + 0.02
SocialNet(**SWE**)	7.06	0.0864 + 0.02
MediaStream(**ES**)	166.17	0.0456 + 0.02
MediaStream(**SWE**)	16.13	0.0432 + 0.02

TABLE II: Comparison of carbon and local operational cost per location, plus the cost of moving the app from ES to SWE.

one, while requiring less memory and similar storage capacity. This performance difference arises because composing and uploading a movie review is more computationally demanding than creating a social media post.

Carbon footprint. Next, we capture the difference of running the applications in the two distinct countries mentioned above. Table II captures the total carbon emissions measured in milligrams of CO2 equivalent (mgC02eq). To calculate this number, we aggregate the total energy consumption during execution, and multiply it with the country's carbon intensity, as suggested in [10]. We observe that running applications in Sweden reduces emissions by an order of magnitude, aligning with the difference in the carbon intensity of the two countries. Finally, we see that the media streaming application generates more than $2\times$ higher carbon emissions than the social network, due to its longer latency, as reported in Table I. Thus, hosting the media streaming application in Sweden will lead to a greater impact on environmental sustainability.

Resource cost. To realize a greener deployment, a company located in Spain needs to expand its infrastructure with additional resources in Sweden, leading to additional resource costs. To quantify those, we use the Amazon EC2 On-Demand Pricing [11] and consider the `t3.large` instance for the social network application and the `t3.medium` instance for the media streaming application, to accommodate the difference in memory footprint. Table II reports the on-demand hourly rate in the `eu-south-2` region for Spain vs. the `eu-north-1` region for Sweden. In addition, we calculate the cost of moving the entire application from Spain to Sweden. by multiplying the storage size with the data transfer cost per GB from Spain to Sweden in AWS [11]. We observe that deploying both applications in Sweden requires a cost similar to deploying them in Spain. Thus, if an application is moved within Europe, then double the budget is needed to acquire similar infrastructure in a different country, because the application needs to run on both locations, to meet performance requirements. Also, there is an additional cost associated with the actual migration, which is proportional to the application's storage needs. Finally, we observe that the cost of deploying the media streaming application in Sweden is half of that of the social network one.

II. Lessons Learned & Discussion

Our findings highlight the trade-off between carbon emissions and cost for two real-world microservice applications, media streaming and social network. For both of the applications that we studied, migration to a greener area has **an order of magnitude reduction in carbon emissions, while it doubles the cost of deployment**. In addition, we observe that the media streaming application generates twice the carbon emissions as the social network, and it has a smaller memory and storage footprint, which translates to reduced migration costs and overheads.

The above observations indicate that a greener deployment is not always affordable for a small or medium-sized national company, since it could double its operational cost. However, a **smart choice** of *which* application to migrate to greener locations can lead to a significant reduction in the absolute number of carbon emissions, in return for reduced migration overheads. Therefore, we conclude that extensive and insightful workload characterization is critical for expanding operations in greener regions by smart offloading of selected applications. Our future work will build upon this insight and design a new system solution that deploys online selective workload migration, and focuses on applications that maximize carbon savings while minimizing cost.

References

[1] A. Lechowicz, N. Christianson, J. Zuo, N. Bashir, M. Hajiesmaili, A. Wierman, and P. Shenoy, "The online pause and resume problem: Optimal algorithms and an application to carbon-aware load shifting," *Proc. ACM Meas. Anal. Comput. Syst.*, vol. 7, no. 3, Dec. 2023. [Online]. Available: https://doi.org/10.1145/3626776

[2] W. A. Hanafy, Q. Liang, N. Bashir, A. Souza, D. Irwin, and P. Shenoy, "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*, ser. ASPLOS '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 479–496. [Online]. Available: https://doi.org/10.1145/3620666.3651374

[3] T. Sukprasert, A. Souza, N. Bashir, D. Irwin, and P. Shenoy, "On the limitations of carbon-aware temporal and spatial workload shifting in the cloud," in *Proceedings of the Nineteenth European Conference on Computer Systems*, ser. EuroSys '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 924–941. [Online]. Available: https://doi.org/10.1145/3627703.3650079

[4] W. A. Hanafy, Q. Liang, N. Bashir, D. Irwin, and P. Shenoy, "Carbonscaler: Leveraging cloud workload elasticity for optimizing carbon-efficiency," *Proc. ACM Meas. Anal. Comput. Syst.*, vol. 7, no. 3, Dec. 2023. [Online]. Available: https://doi.org/10.1145/3626788

[5] V. Gsteiger, P. H. D. Long, Y. J. Sun, P. Javanrood, and M. Shahrad, "Caribou: Fine-grained geospatial shifting of serverless applications for sustainability," in *The 30th ACM Symposium on Operating Systems Principles (SOSP'24). ACM*, 2024.

[6] N. Bashir *et al.*, "The sunk carbon fallacy: Rethinking carbon footprint metrics for effective carbon-aware scheduling," in *Proceedings of the 2024 ACM Symposium on Cloud Computing*, 2024, pp. 542–551.

[7] "Electricity maps," https://app.electricitymaps.com/.

[8] Y. Gan *et al.*, "An open-source benchmark suite for microservices and their hardware-software implications for cloud & edge systems," in *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*, ser. ASPLOS '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 3–18. [Online]. Available: https://doi.org/10.1145/3297858.3304013

[9] "Deathstar benchmark," https://github.com/delimitrou/DeathStarBench.

[10] L. Lannelongue, J. Grealey, and M. Inouye, "Green algorithms: quantifying the carbon footprint of computation," *Advanced science*, vol. 8, no. 12, p. 2100707, 2021.

[11] "Aws pricing on-demand," https://aws.amazon.com/ec2/pricing/on-demand/.