TCSS 562 Fall 2024

Cloud Research Paper Presentation
5 December 2024

FootPrinter: Quantifying Data Center Carbon Footprint

Authors: Dante Niewenhuis, Sacheendra Talluri, Alexandru Iosup, and Tiziano De Matteis

Institution: Vrije Universiteit Amsterdam (Amsterdam, Netherlands)

Review by: Carla Peterson and Viktoria Dolojan

Talk Outline

What is the problem?

Data Centers:

- Major carbon emitters
- Complex footprint quantification
  - Diverse energy sources
  - Fluctuating utilization patterns
  - Indirect emissions

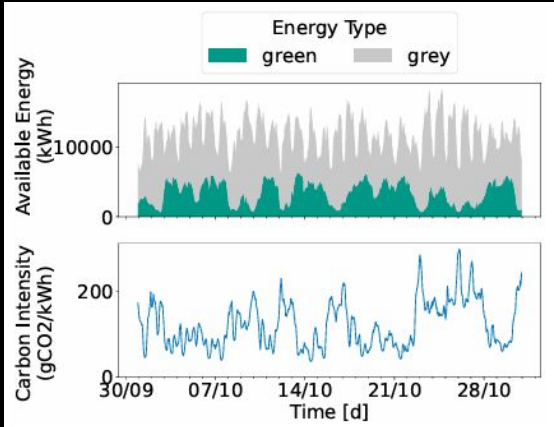


Figure 1: The energy mix and carbon intensity of the energy grid in the Netherlands during the month of October 2023 from the ENTSO-E Transparency Platform[1]. The top graph shows the energy mix during the month into green and non-green energy. The bottom graph shows the resulting carbon intensity of the grid.

Why is it a problem?

Rising Energy Demand:
Data center emissions will escalate alongside their expanding role in the economy

Impact on Climate Goals:
Stakeholders put global climate goals at risk
Delay progress toward reducing greenhouse gas emissions

Inefficiency and Lack of Transparency:
Current practices can lead to inefficient energy use
Poor understanding of how to improve sustainability

Why are we interested in the problem?

Addressing Climate Change:
Enhancing data center energy efficiency can reduce emissions while enabling technological advancements

Creating Standardized Benchmarks:
Reliable metrics for carbon footprints allow organizations to evaluate and improve their sustainability efforts

Driving Policy and Industry Alignment:
Standardized tools can help stakeholders achieve shared sustainability goals

Related Work

Specialized Simulators

CloudSim:
Models and simulates cloud computing environments

iFogSim:
Models and evaluates the performance of Internet of Things (IoT) and fog computing environments

WorkflowSim:
Models the execution of scientific workflows in cloud computing environments

Green Algorithms:
https://www.green-algorithms.org/

Analyzes and optimizes the energy consumption of cloud computing algorithms

Missing Elements

Indirect Emissions Modeling:

Existing tools fail to account for lifecycle emissions

Dynamic Energy Grids:

Simulators often oversimplify energy source diversity

Comprehensive Frameworks:

No single simulator integrates edge, fog, and cloud computing with lifecycle carbon footprint assessments and real-time renewable energy modeling

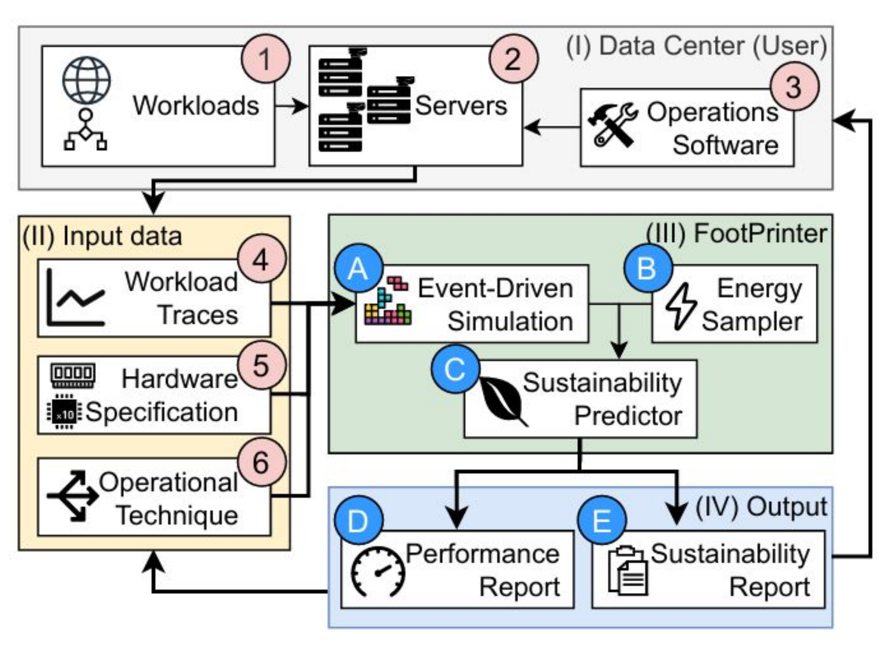
What is being proposed?

FootPrinter

Designed to bridge previous gaps

Offers enhanced adaptability and functionality

Versatile for assessing carbon footprints across a wide range of energy sources

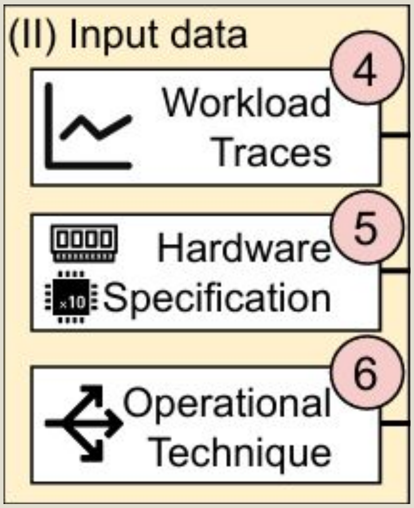


Input Data

Workload Traces:
Job logs
Hardware needs
Demand
Timing

Hardware Specification:
Server setup
Location
Energy grid link

Operational Techniques:
Scheduling
Resource allocation

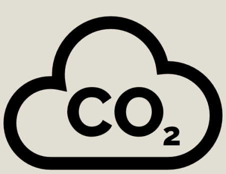


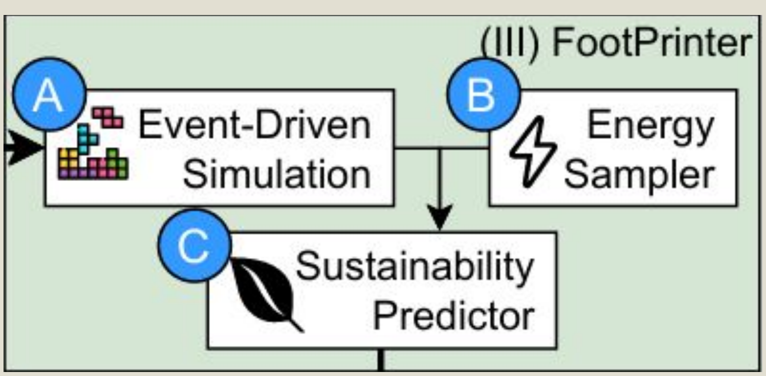
FootPrinter Process

Event-Driven Simulator:
Replay
Sample energy/performance

Energy Sampler:
Carbon intensity from ENTSO-E data

Sustainability Predictor:
Carbon/energy metrics
Footprint





Output Data

Performance Report:
Job time
CPU usage
Efficiency

Sustainability Report:
Energy use
Carbon emissions
Trends

Contributions

FootPrinter: First tool to simulate operational carbon footprint of data centers

Open-Source: Extensible for hardware, environmental factors, and embodied emissions

Energy Agnostic: FootPrinter does not depend on the energy source

Findings

Use Cases: Demonstrates carbon impact of design and location choices

Insights: Actionable guidance for sustainable data center operations

Broad Access: Supports varied trace granularities for diverse operators

- Power Usage Effectiveness (PUE)
  - E_T energy used by the data center
  - E_{IT} energy used by the IT components
  - Optimal data center has PUE = 1.0 no energy required for redundant tasks
  - Google has PUE ~ 1.1
  - PUE of data centers has increased
    - Rebound effect - as prices to perform tasks decrease, # of tasks increase
    - PUE already highly optimized & difficult to further optimize
  - Does not include energy efficiency of applications & workloads
  - Ignores type of energy used (e.g. solar, wind, coal, nuclear)

$$PUE = \frac{E_T}{E_{IT}}$$

- Carbon Intensity
  - Amount of carbon emitted per unit energy
  - Carbon intensity of the grid
    - CI_s carbon intensity of energy source s
    - E_s / E_g share of energy that s contributes to the grid
    - S set of all available energy sources

$$CI_g = \sum_{s \in S} CI_s \frac{E_s}{E_g}$$

- Operational Footprint
  - Carbon emitted when system is running
  - CI_d carbon intensity of the data center in gCO2/kWh
    - Assume $CI_d = CI_g$
  - E_{op} operational energy of the data center in kWh

$$C_{op} = CI_d E_{op}$$

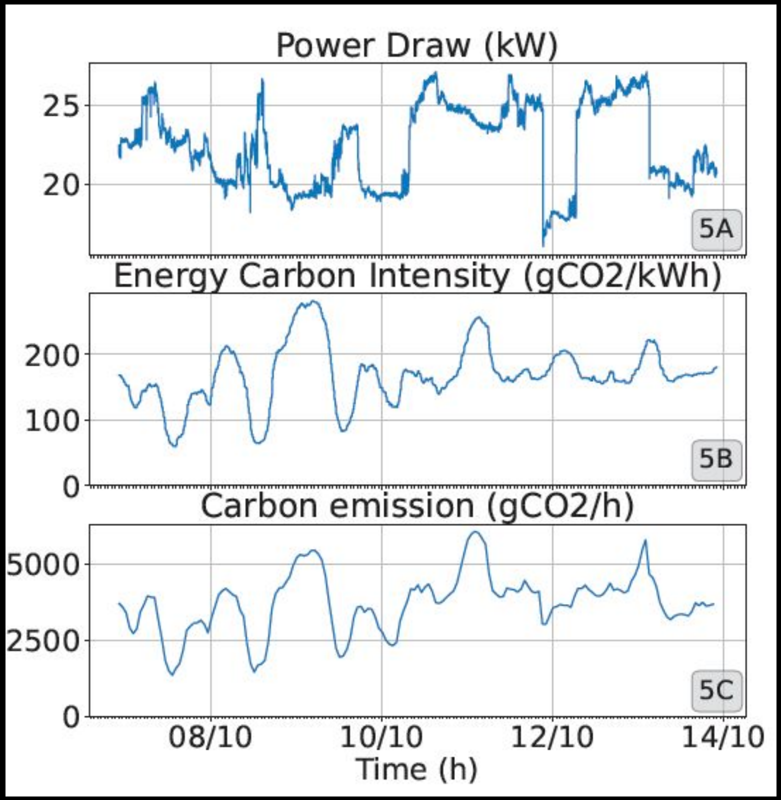
- UC-Footprint
  - Operational carbon footprint
  - Essential to evaluating a data center’s effectiveness
  - Requires knowledge about both the energy usage and the carbon intensity of the used energy sources
- UC-Location
  - Selecting a location
  - Can have a large impact on the carbon footprint due to what’s available for energy sources
  - Where’s the right location?
- UC-Hardware (not evaluated)
  - Selecting hardware upgrades
  - Responsible for making the right choices for hardware upgrades
  - Must understand impact of hardware changes

Dataset Input to FootPrinter

- Data center: SURF Lisa cluster in Netherlands
  - 277 physical machines
  - 7,850 jobs over 7 days
  - Job duration: <1 hr to several days
  - CPU demand sampled @ 30-sec interval
- FootPrinter simulates trace on Intel Core I7-8750 laptop in 10 secs

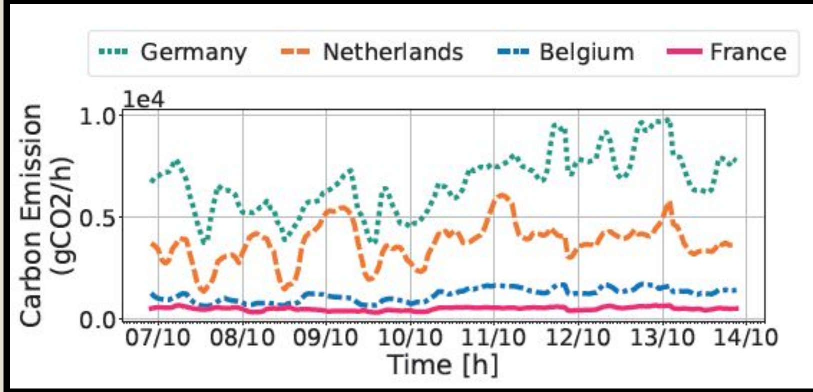
FootPrinter Simulation Results

- 5A: FootPrinter determined power draw of entire data center
- 5B: Carbon intensity of the grid sampled from ENTSO-E
- 5C: Carbon emission during the workload. Influenced by carbon intensity.



FootPrinter Simulation Results

- Compare impact of data center's location
- SURF Lisa data center workload trace simulated in different locations
- France & Belgium have better carbon footprints
  - Source nearly ½ of energy from nuclear power plants
- Germany & Netherlands more carbon intensive energy sources (e.g. coal)



SURF Lisa data center workload trace simulated using FootPrinter in other locations

- Compared FootPrinter's simulated power draw to ground truth power
  - No data source provided
- Mean Absolute Percentage Error (MAPE)
  - MAPE total error: 3.15%
  - Underestimation error: 3.19%
  - Overestimation error: 2.93%

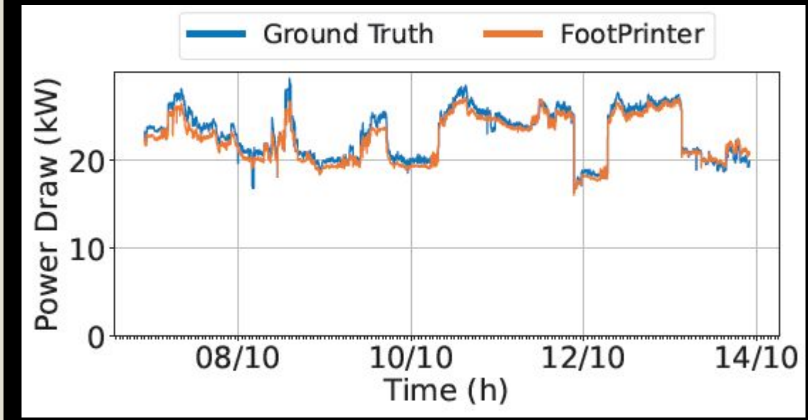
$$MAPE[\%] = \frac{1}{n}\sum_{t=0}^{n} |\frac{P_t - P'_t}{P_t}| \times 100$$

- Normalized Absolute Difference (NAD)
  - NAD total error: 3.17%
  - Underestimation error: 3.22%
  - Overestimation error: 2.83%

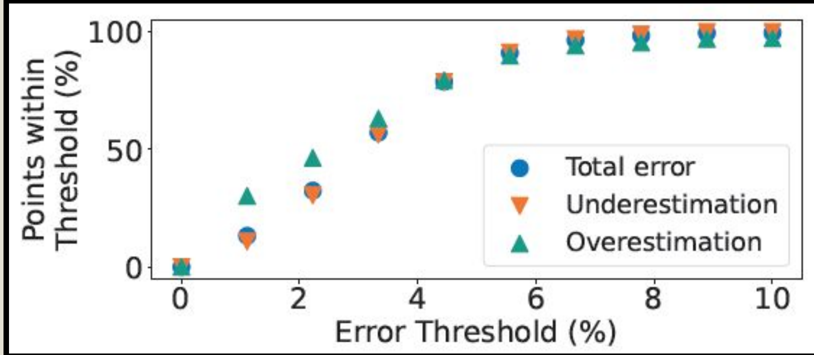
$$NAD[\%] = \frac{\sum_{t=0}^{n} |P_t - P'_t|}{\sum_{t=0}^{n} P_t} \times 100$$

In which P_t and P'_t are the actual and simulated power draw at sample t and n is the number of samples. Comparing FootPrinter

- Percentage of time points with an error less the specified threshold



Power Draw Comparison



Distribution of the Error of Samples

- FootPrinter is the first simulation tool to determine operational carbon footprint of a data center
- Open-source and available for extension for additional tools
- Simulates power draw with MAPE < 3.15% & NAD < 3.17%

- Addresses the carbon footprint generated by data centers through predictive simulation
- FootPrinter can simulate the operational carbon footprint of a data center regardless of the type of energy source
- Low total MAPE (3.15%) and NAD (3.17%)

- Focus seems to be within the European region
  - Would like to see more of a US focus for more complex examples
- Small dataset for initial evaluation and validation
  - Only one data center, SURF Lisa, was used for the experiment and validation
  - Additional data centers for experiment and validation would create more confidence in FootPrinter
- No reference for ground truth power draw data
  - Creates very little confidence in their validation efforts without a data source

- FootPrinter needs additional evaluation and validation with additional workload traces from data centers
- No discussion on the results of the use cases
  - Results were stated, but no discussion on why the results appeared as they were (e.g. dip in power draw)
- No discussion of threshold or listing a specific threshold for the distribution of errors graph

- Supporting hardware upgrades & impact to performance and carbon footprint
- Future work:
  - Additional elements that influence energy usage: temperature & humidity
  - Extend to incorporate embodied carbon emissions

THANK YOU!

Questions?