UMassAmherst | Manning College of Information & Computer Sciences

CarbonEdge: Leveraging Mesoscale Spatial Carbon-Intensity Variations for Low Carbon Edge Computing

Li Wu[1], Walid A. Hanafy[1], Abel Souza[2], Khai Nguyen[1], Jan Harkes[3], David Irwin[1], Mahadev Satyanarayanan[3], Prashant Shenoy[1]

[1]University of Massachusetts Amherst
[2]University of California, Santa Cruz
[3]Carnegie Mellon University

July 22, 2025

NSF | EXPEDITIONS: CoDEC

The 34th ACM International Symposium on High-Performance Parallel and Distributed Computing

Carbon Emissions of Computing

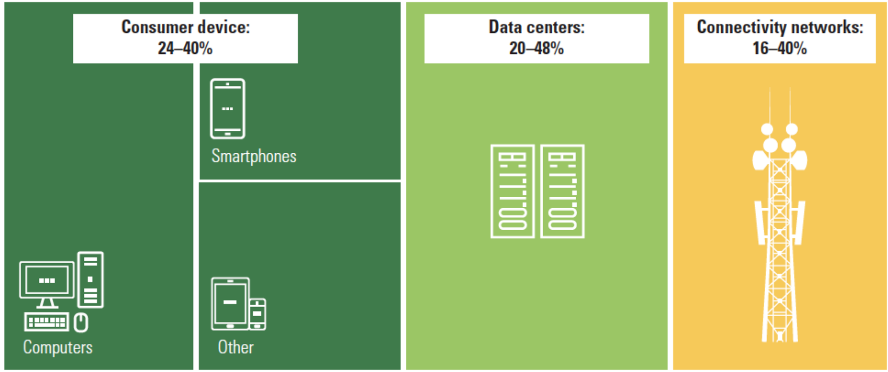
Climate change impacts

Wildfire

Drought

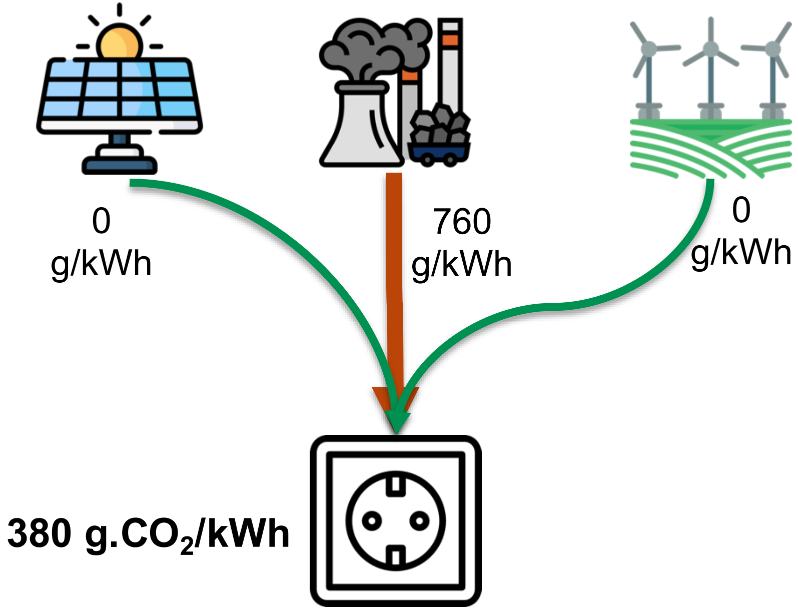
Tropical Cyclones: Wind and Rain

1.5 - 4% of Global Carbon Emissions is from **ICT**.



Source: Adapted from WIK-Consult and Ramboll (2021) to include estimates by Minges, Mudgal, and Decoster (forthcoming) based on analysis of reported emissions by more than 150 international digital companies.

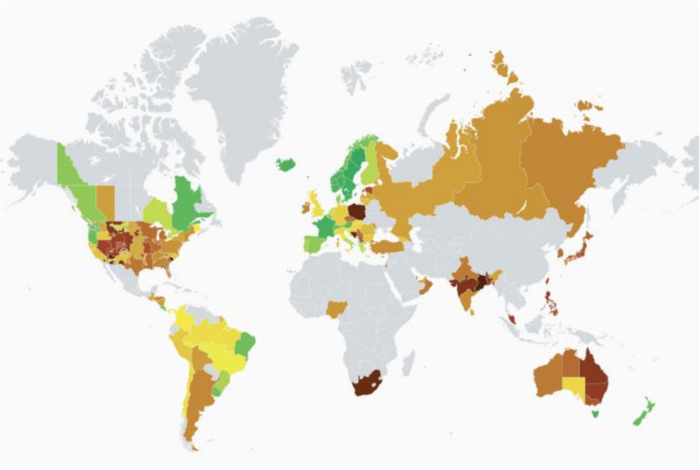
Carbon Intensity and Carbon Emissions



Carbon Intensity

X **Energy Consumption** = **Carbon Emissions**

Carbon Savings across Cloud Data Centers



Energy demand and supply mix changes across space.

Spatial shifting saves XX carbon footprint, while incurring YY latency.

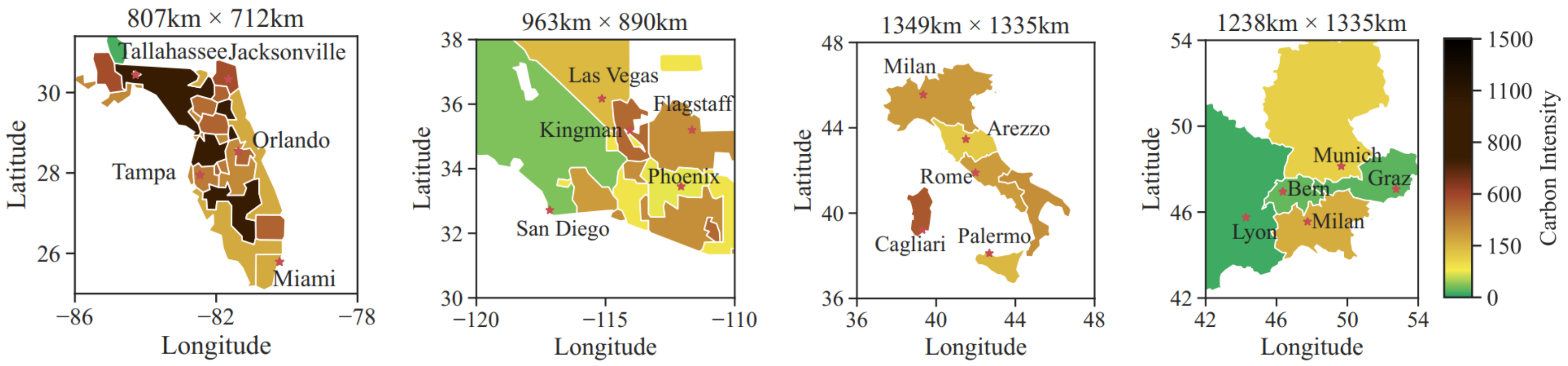
How about Edge Data Centers?

Is it possible to save carbon while meeting low-latency requirements?

How much does carbon intensity vary within mesoscale regions?

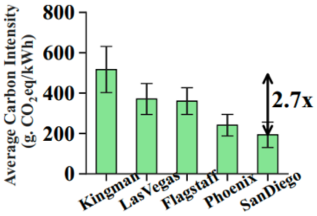
How prevalent are these types of mesoscale variations ?

Mesoscale Carbon Analysis

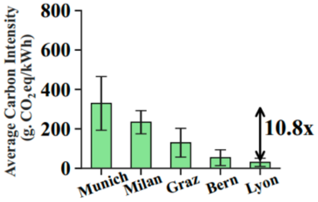


(a) Florida (b) West US (c) Italy (d) Central EU

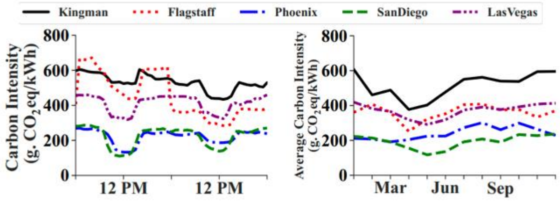
Carbon Intensity Variations and Latency



(a) West US



(b) Central EU



(a) Two-day (Dec 25-27) (b) Year-long (2023)

Table 1: One-way network latency (ms).

(a) Florida

Location	Miami	Orlando	Tampa	Tallah.
Jacksonville	3.64	5.32	6.86	3.42
Miami	-	4.5	3.37	7.2
Orlando		-	1.86	4.35
Tampa			-	4.14
Tallahassee				-

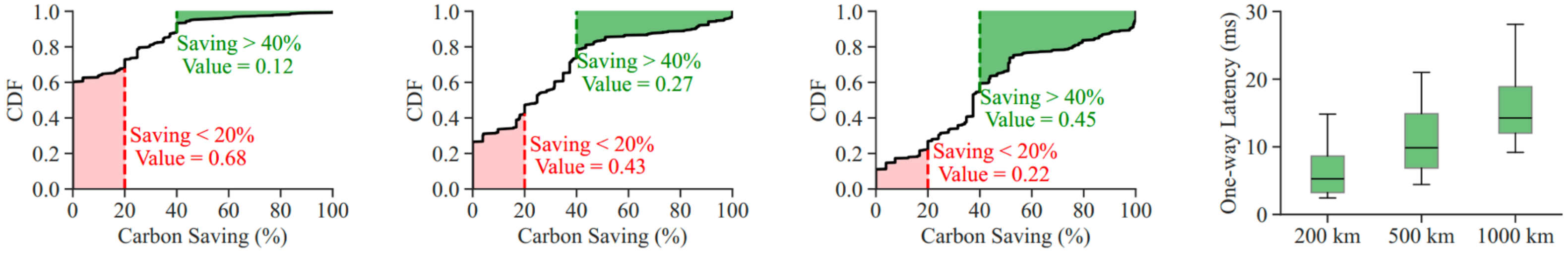
(b) Central EU

Location	Graz	Lyon	Milan	Munich
Bern, CH	8.78	6.28	6.45	3.985
Graz, AT	-	16.22	11.98	8.36
Lyon, FR		-	9.34	8.82
Milan, IT			-	8.65
Munich, DE				-

Significant differences in the carbon intensity of electricity at mesoscale distances with XX ms network latency.

Mesoscale Analysis across Continents

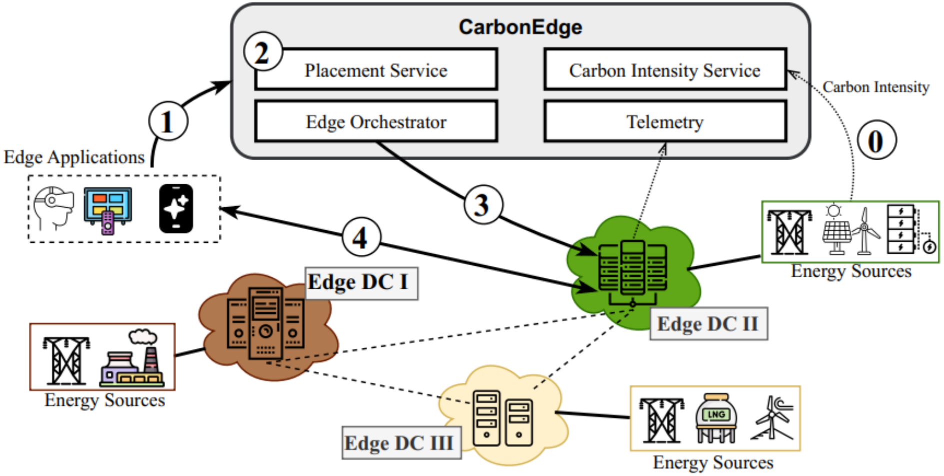
- 496 Akamai edge data center



(a) D = 200 km **(b) D = 500 km** **(c) D = 1000 km** **(d) Radius-Latency**

More than 78% of the edge locations in Europe and North America see carbon intensity differences exceeding 20% within a radius of 1000 km.

CarbonEdge Overview



CarbonEdge: Carbon-aware framework for edge computing

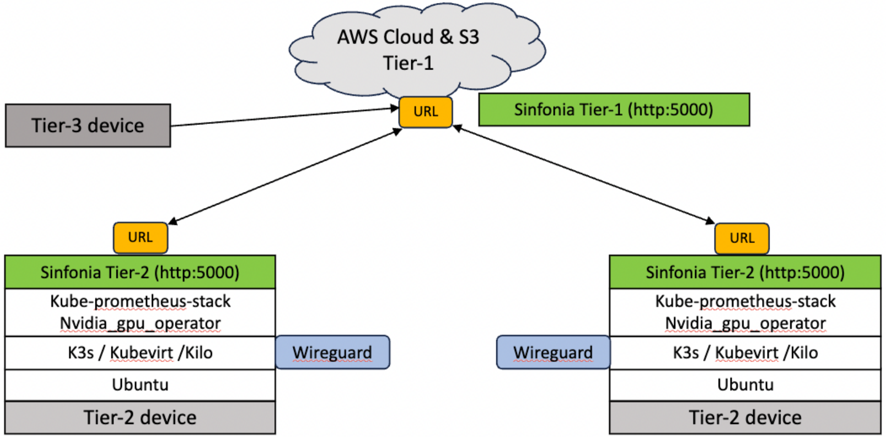
Carbon-aware Edge Placement

- Incremental carbon-aware placement
- Optimize the application placement and server activation
- Consider the heterogeneity of servers
- Carbon emissions from application operation and server activation
- ILP for incremental application placement

$$f = \underbrace{\sum_i \sum_j x_{ij} \cdot E_{ij} \cdot \bar{I}_j}_{\text{Application operation}} + \underbrace{\sum_j (y_j - y_j^{curr}) \cdot B_j \cdot \bar{I}_j}_{\text{Server activation}}$$

CarbonEdge Implementation

- Sinfonia
- Telemetry Service
  - Power Monitoring (RPAL, DMPG)
  - Carbon Intensity (Electricity Map)
  - Carbon Monitoring
  - End-to-end latency
- Profiling Service
- Placement Service

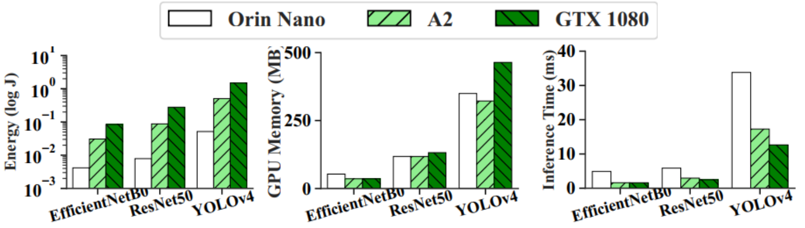


Sinfonia

Experimental Setup

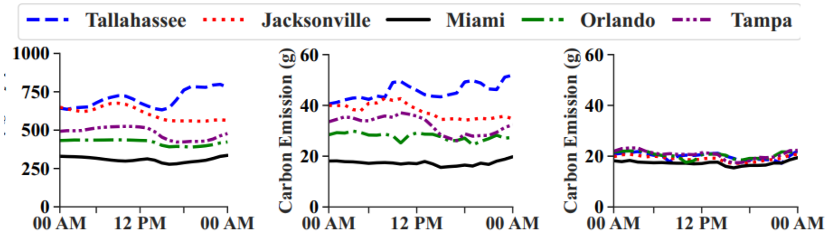
- Real World Traces
  - Carbon Intensity Traces from Electricity Map
  - Latency Traces from WonderNetwork
  - Edge Workloads (CPU and GPU)
  - Edge Data Centers from Akamai CDN
- Edge Testbed
  - 5 edge data centers, each with one server
  - Dell PowerEdge R630 with an NVIDIA A2 GPU
  - Latency injection with tc
- Three Baselinse
  - Latency-aware
  - Energy-aware
  - Intensity-aware

- Model Profiling
  - Energy, GPU memory, and inference time
  - 3 GPUs: A2, Orin Nano, GTX 1080

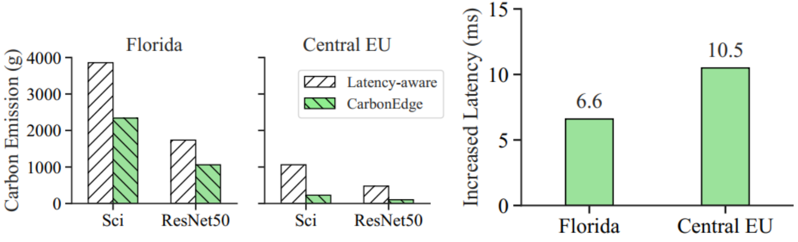


(a) Energy (b) GPU Memory (c) Inference Time

Mesoscale Regional Edge Deployment



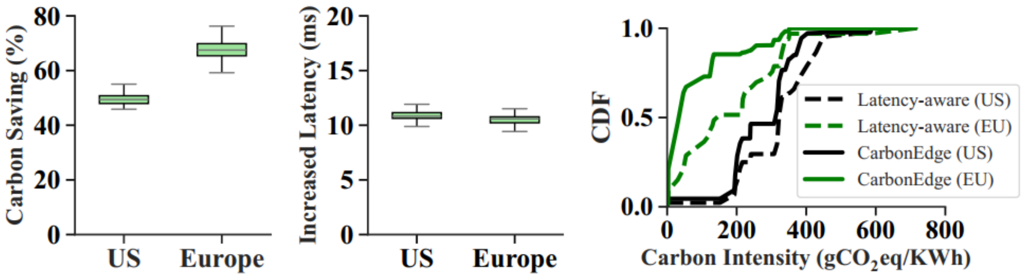
(a) Carbon Intensity (b) Latency-aware (c) CarbonEdge



(a) Carbon Emissions (b) Latency Increases

In mesoscale edge settings, CarbonEdge can highly optimize the carbon emissions resulting in 39.4% and 78.7% carbon savings for Florida and Central EU, respectively.

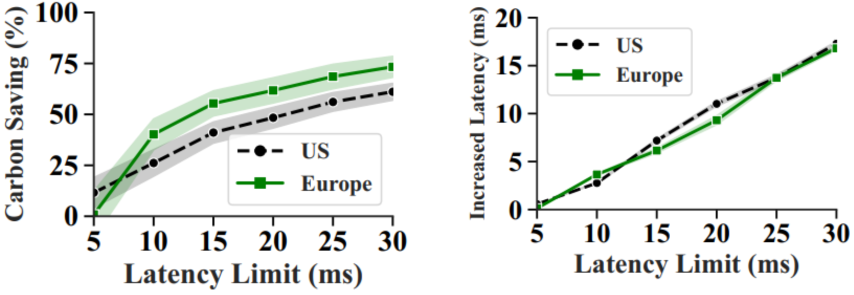
Continental-scale CDN Edge Deployment



(a) Carbon (b) Latency (c) Load Distribution

By shifting the demand towards low carbon zones, CarbonEdge decreases carbon emissions by 49.5% and 67.8% for the US and Europe, respectively, while increasing the round-trip latency by less than 11 ms.

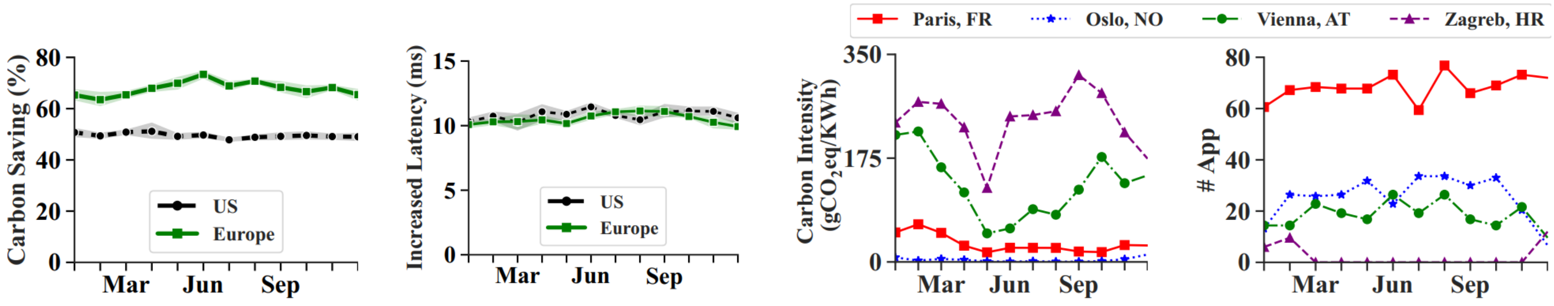
Impact of Latency Tolerance



(a) Carbon Savings (b) Latency Increases

For a 10 ms increase in latency, CarbonEdge derives 28% and 44.8% carbon savings in the US and Europe, respectively.

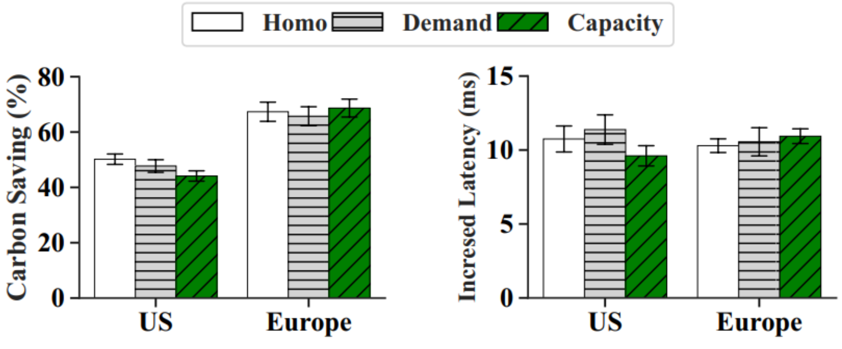
Impact of Seasonality



(a) Carbon Savings (b) Latency Increases (c) Carbon Intensity (d) Placements

The seasons' changes in carbon intensity highly affect the carbon savings that change by up to 10% across months. The intertwined relations between regions change across seasons, resulting in up to 3× change in resource allocation.

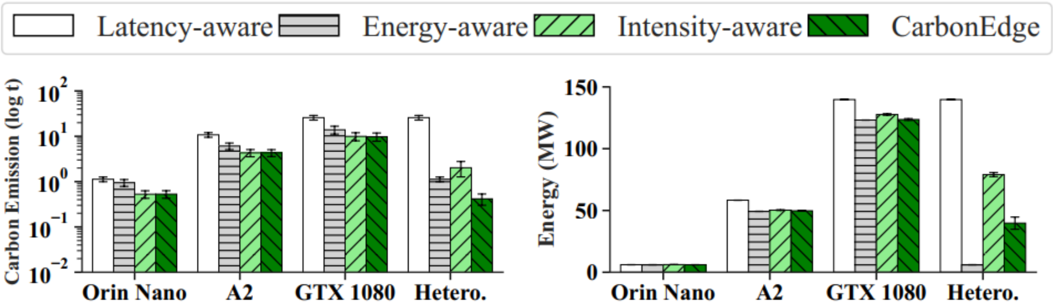
Impact of Demand and Capacity



(a) Carbon Savings (b) Increased Latency

Changes in demand and capacity can impact carbon savings based on the carbon intensity of their origin.

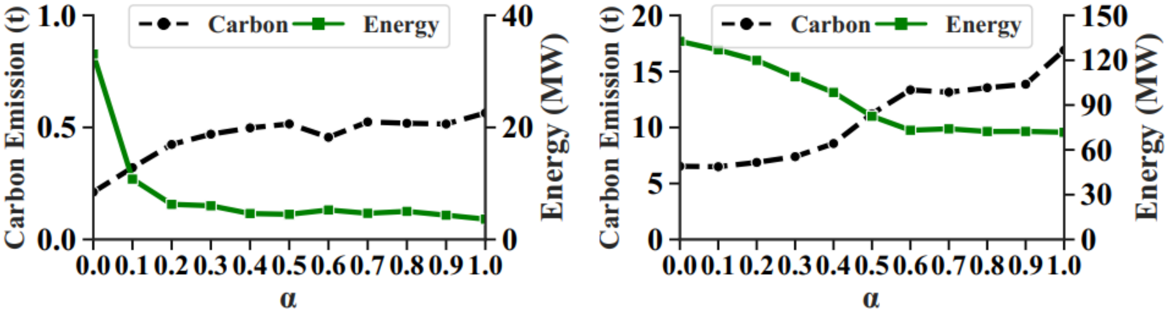
Impact of Heterogeneity



(a) Carbon Emissions (log t) (b) Energy Consumption

By interplaying the differences in energy efficiency, carbon intensity, and processing speed, CarbonEdge can reduce carbon emissions by 98%, 79%, and 63% compared to the Latency-aware, Intensity-aware, and Energy-aware baselines, respectively.

Carbon-Energy Trade-off

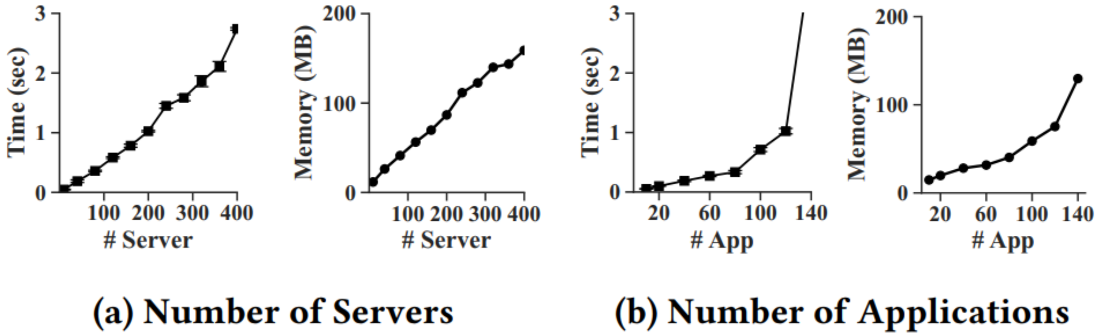


(a) Low Utilization

(b) High Utilization

The inherent carbon-energy trade-off is pronounced in heterogeneous edge settings. By augmenting the objective function with energy-awareness, CarbonEdge can maintain 97.5% of its carbon savings while decreasing energy consumption by 67%.

System Overhead



Complete computation within 3 seconds with 400 servers and 140 applications, consuming less than 200 MB memory.

Conclusion

- **Mesoscale Carbon Analysis** reveals significant spatial variations in carbon intensity across edge data centers.
- **CarbonEdge**: a carbon-aware framework with an incremental application placment optimization algorithm.
- **Extensive evaluation** on a mesoscale regional edge testbed and continental-scale edge simulations demonstrates its effectiveness.

Carbon savings outweigh latency overhead in edge computing!

UMassAmherst | Manning College of Information & Computer Sciences

COMPUTING FOR THE COMMON GOOD

Li Wu
liwu@cs.umass.edu