Strategies for operating and sizing low-carbon cloud data centers

Miguel Felipe SILVA VASCONCELOS[1,2]

December 20, 2023

Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, Grenoble, France[1]
School of Sciences, Arts, and Humanities, University of São Paulo, Brazil[2]

Cloud computing

- Provide computational resources on demand
- Support the marjority of services and applications we use (from social networks to health and banking applications)
- 1% of the worlds electricity demand
- 1% of the CO_2 emissions from electricity

Dealing with environmental impact

- Cloud providers incorporating renewable energy (Google reported an annual average of x% up to y%)
- Improvements in efficiency in hardware and software: from 2010 to 2018 6% in workload and only 6% in energy consumption
- Uncertainty with future projects: may double with end of Moore's law and rise of IOT applications
- Energy consumption, is not only source of environmental impact. GHG protocol has three scopes
  - Direct emissions
  - Emissions from power consumption
  - Other emissions - 76% of total emissions from cloud data centers.

Strategies for operating and sizing low-carbon cloud DCs

- Low-carbon cloud data centers ?
  - Considering the 3 scopes of the GHG protocol
- Which strategies?
  - Carbon-responsive strategies follow-the-renewables and sizing for the IT and renewable infrastructure

My research interests

Scheduling the workload Usage of follow-the-renewables:

- Allocates/Migrates the workload to the data centers (DCs) that have more renewable (green) power available
- Migrating the workload among different DCs generates extra computations proportional to the duration of the migration

Impact of follow-the-renewables[1]

- Baselines that do not consider network:
  - Migration time > 100 times longer
  - Wasted energy could have powered one of the DCs for 44 hours
- Proposed solution:
  - Migration algorithm that considers network bandwidth, topology, and the history of usage of the links
  - No network congestion and same or lower brown energy consumption

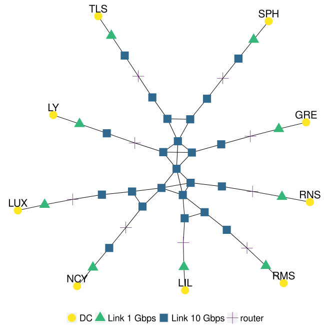

Figure 1: DCs and how they are connected in the network.

[1] Miguel Felipe Silva Vasconcelos, Daniel Cordeiro, and Fanny Dufossé. "Indirect Network Impact on the Energy Consumption in Multi-clouds for Follow-the-renewables Approaches." In: *11th International Conference on Smart Cities and Green ICT Systems*. 2022.

My research interests

Sizing the renewables infrastructures:

- Compute the area of solar panels (PVs) in m^2 and batteries capacity in Wh

Collaboration with researchers from the DATAZERO 2 project:
Jean-Marc Nicod, Georges Da Costa, Veronika Rehn-Sonigo

Sizing renewable infrastructures

Data centers:

- Infrastructure already built (servers, network)
- Homogeneous (regarding CPU cores)
- Server power consumption: idle and dynamic
- Intra network power consumption: static
- Specific Power Usage Effectiveness (PUE) for each DC

Figure 2: Selected data centers location (inspired from Microsoft Azure)

Workload:

- All tasks must be scheduled and executed on time
- Batch tasks that can be executed in any of the DCs
- No migration
- Task execution cannot be delayed

Renewables infrastructure

- Batteries charge and discharge efficiency, Maximum Depth of Discharge
- PV panels efficiency
- Carbon emissions from manufacturing (PV: 250 kg CO_2 eq per m², bat: 59 kg CO_2 eq per kWh)
- Lifetime (PV: 30 years, bat: 10 years)

Local electricity grid

- The energy mix is different at each location
- May have the presence of renewables or low carbon-intensive sources

Table 1: Emissions (in g CO_2-eq/kWh) for using the regular grid. Source for grid emissions: electricityMap, climate-transparency.org.

Location	Emissions
Johannesburg	900.6
Pune	702.8
Canberra	667.0
Dubai	530.0
Singapore	495.0
Seoul	415.6
Virginia	342.8
São Paulo	61.7
Paris	52.6

Proposed solution

Linear program formulation to minimize the carbon emissions from the cloud federation operation (timespan of 1 year) [2]

- Scheduling and dimensioning modeled as single problem
  - Allocate workload to other DC or increase the battery capacity or PV area?
- Only real variables
  - Optimal solution in polynomial time: 394264 variables, solved in less than 1 minute with Gurobi

[2] M. Vasconcelos, D. Cordeiro, G. Da Costa, F. Dufossé, J.-M. Nicod, and V. Rehn-Sonigo, "Optimal sizing of a globally distributed low carbon cloud federation". *In: 2023 23nd IEEE International Symposium on Cluster, Cloud and Internet Computing (CCGrid), Bengaluru, India, 2023.*

Linear Program summary

Obj. function: Minimize the DC's operation CO_2 emissions (1 year)

- CO_2 comes from: grid power, manufacturing PV and batteries

Linear Program summary

Obj. function: Minimize the DC's operation CO_2 emissions (1 year)

- CO_2 comes from: grid power, manufacturing PV and batteries

Input

- Power consumption of servers and network equipment
- 1 year of workload (Google trace)
- Renewable infrastructure specs (efficiency, manufacturing CO_2)
- For each DC:
  - Solar irradiation (1 year)
  - CO_2 from electricity grid
  - PUE
  - CPU cores number

Linear Program summary

Obj. function: Minimize the DC's operation CO_2 emissions (1 year)

- CO_2 comes from: grid power, manufacturing PV and batteries

Input

- Power consumption of servers and network equipment
- 1 year of workload (Google trace)
- Renewable infrastructure specs (efficiency, manufacturing CO_2)
- For each DC:
  - Solar irradiation (1 year)
  - CO_2 from electricity grid
  - PUE
  - CPU cores number

Output

- PVs area (m^2)
- Batteries capacity (kWh)
- Carbon emissions
- Schedule of the workload

Results

Table 2: Total emissions for the different scenarios.

Scenarios	Emissions (t CO_2-eq)
Electrical grid	201 211.3
PV and batteries	42 370.6
PV, batteries, and grid	29 600.6

Reductions on carbon emissions:

- Grid vs DC renewable infra : $\simeq 5$ times
- Grid vs hybrid configuration (DC renewables and grid) : $\simeq 6$ times

Results

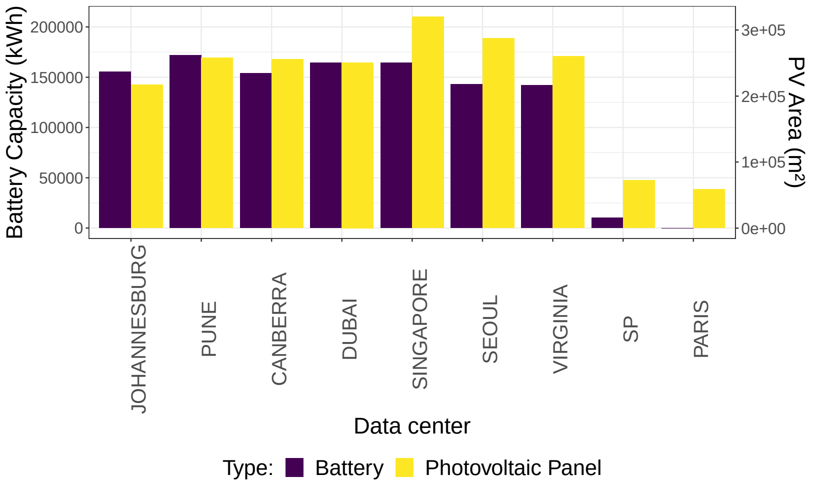

Figure 3: Optimal result for the area of PV panels and capacity of the batteries.

Results

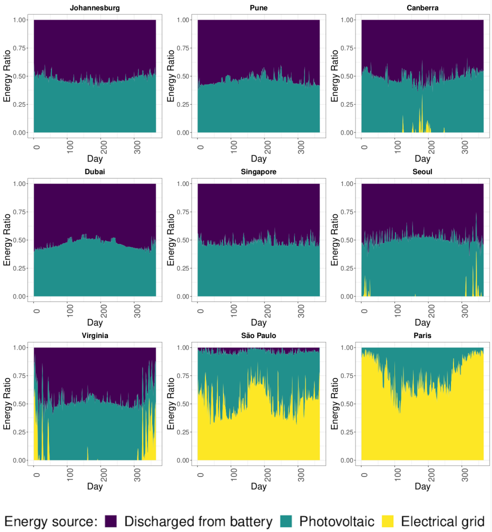

Figure 4: Composition of the DCs' daily energy consumption throughout the year considering the different sources of energy

Sizing for the long term:

Evaluate other strategies to further reduce the emissions and for the long term:

- Consider the emissions from the life cycle
- How much including wind power could reduce the emissions?
- Impact of other scheduling algorithms in the total emissions (for example, allowing to delay some tasks)
- How expensive is it to reduce the carbon emissions of the cloud operation (monetary costs in dollars)
- Dimensioning of IT infrastructure (servers considering the footprint of manufacturing, new generations that are more efficient)

Impact of adding wind turbines (WT)

- Can further reduce 34% the carbon emissions in comparison to only using PVs and batteries.
- However, requires larger land area (1 to 3 WT per km^2)

Table 3: Computed number of WT for each location.

Location	Number of WT
Johannesburg	59
Pune	26
Canberra	67
Dubai	79
Singapore	37
Seoul	109
Virginia	39
São Paulo	87
Paris	22

Impact of adding wind turbines (WT)

- Can further reduce 34% the carbon emissions in comparison to only using PVs and batteries.
- However, requires larger land area (1 to 3 WT per km²)

Table 4: Capacity Factor (in %) for solar panels and wind turbines at each location.

Location	PV	WT
Johannesburg	25.55	12.96
Pune	24.26	10.04
Canberra	22.08	12.97
Dubai	25.28	13.98
Singapore	17.68	8.58
Seoul	18.81	9.41
Virginia	19.83	14.68
São Paulo	21.74	10.06
Paris	15.37	23.51

Flexibility in the scheduling

What is the impact in carbon emissions of delaying α percent of the jobs up to β time slots (1h per time slot) ?

Flexibility in the scheduling

What is the impact in carbon emissions of delaying α percent of the jobs up to β time slots (1h per time slot) ?

Table 5: Reductions in total carbon emissions (%) in comparison to the scenario where it is not possible to delay the workload.

α \ β	1	24	48	72	96	120	144	168
10	0.46	3.14	3.48	3.66	3.76	3.81	3.85	3.85
20	0.84	3.85	4.11	4.21	4.21	4.21	4.22	4.22
30	1.15	4.07	4.25	4.25	4.26	4.26	4.27	4.27
40	1.42	4.15	4.25	4.26	4.27	4.28	4.28	4.29
50	1.65	4.22	4.26	4.27	4.28	4.29	4.3	4.3

Costs ($) of reducing the environmental impact

- Considered the Levelized Cost of Energy
- Grid price is half at off-peak times (10 pm to 8 am)
- Battery price: 0.20 dolars per kWh of electricity delivered

Table 6: Price of different sources of energy (USD per kWh) at each location.

Location	Grid	PV	WT
Johannesburg	0.074	0.0385	0.1984
Pune	0.104	0.0406	0.2610
Canberra	0.331	0.0445	0.1993
Dubai	0.101	0.0390	0.1863
Singapore	0.272	0.0557	0.3009
Seoul	0.092	0.0525	0.2735
Virginia	0.150	0.0498	0.1760
São Paulo	0.144	0.0453	0.2572
Paris	0.340	0.0643	0.1098

Costs ($) of reducing the environmental impact

Table 7: Total costs (millions of $) and carbon emissions ($t$ CO_2-eq) for the different scenarios.

Scenario	Millions of $	t CO_2-eq
Min. cost (PV + WT + Bat + grid)	36.7	141 001.67
Min. cost (PV + Bat + grid)	38.3	145 193.64
Min. cost (WT + Bat + grid)	51.1	204 568.07
Min. CO_2 (PV + Bat + grid)	72.8	37 828.49
Min. CO_2 (WT + Bat + grid)	277.5	58 263.38
Min. CO_2 (PV +WT+ Bat + grid)	117.3	24 977.89
Only grid	56.7	222 876.62

Sizing the IT part

- Workload growth over time
- New hardware generations that may be more power-efficient
- CO_2 from manufacturing the servers

Sizing the IT part

Scenarios:

- Use the current infrastructure at maximum, and only buy new servers to meet the workload demand
- Can manufacture and replace old servers to minimize the total carbon emissions

Settings:

Table 8: Servers specifications for different generations.

Year	CPU	Cores	Pidle	Pcore	kg CO_2-eq
< 2016	Intel Xeon E5-2660 v2	20	52	7.5	-
2017	Intel Xeon Platinum 8180	56	48.9	6.68	578.6

- Workload 22% increase per year (CISCO)
- Server life time of 4 years

Sizing the IT part

Scenarios:

1. Use the current infrastructure at maximum, and only buy new servers to meet the workload demand
2. Can manufacture and replace old servers to minimize the total carbon emissions

Results:

- Scenario 2 had a 10% lower carbon footprint, and manufactured 8% more servers.

Sizing the IT part

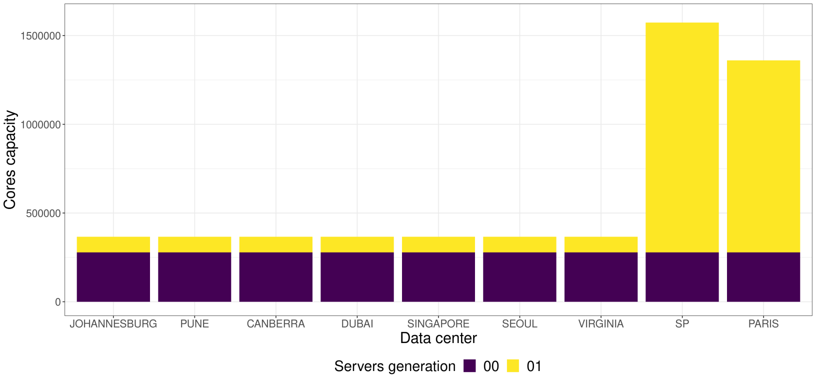

Figure 5: Sizing for the scenario 1 - Only add new servers.

Sizing the IT part

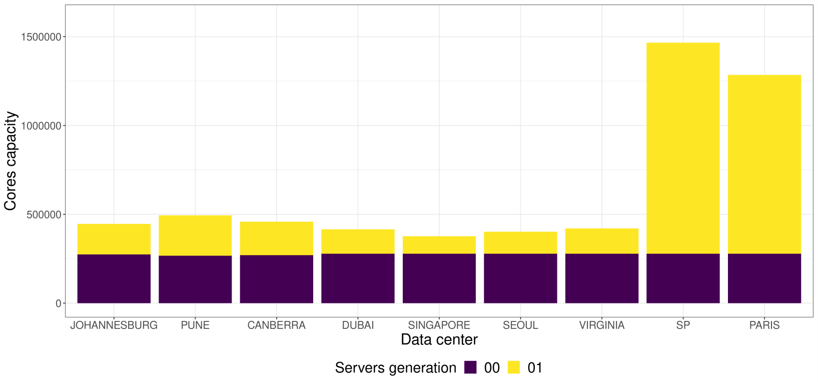

Figure 6: Sizing for the scenario 2 - Add new servers and replace old ones.

Sizing the IT part

Scenarios:

- Decision made year by year (greedy approach)
- Optimal solution (all info is know in advance)

Settings:

Table 9: Servers specifications for different generations.

Year	CPU	Cores	Pidle	Pcore	kg CO_2-eq
< 2016	Intel Xeon E5-2660 v2	20	52	7.5	-
2017, 2018	Intel Xeon Platinum 8180	56	48.9	6.68	578.6
2019, 2020	AMD EPYC 7742	64	66.1	2.71	587.2
2021	AMD EPYC 7763	128	75.6	3	590.3

Sizing the IT part

The optimal solution emits 13.4% less CO_2.

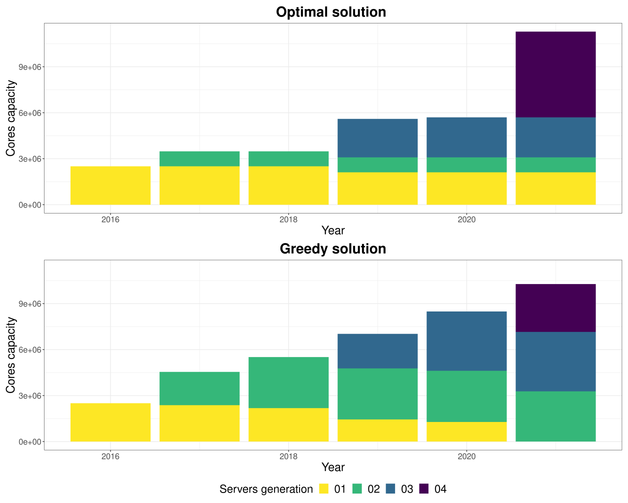

Figure 7: Comparison between the optimal and greedy approaches.

Future work

- What could be learned from the optimal solution of adding/replacing servers ?
- Costs of the servers
- Degradation of the renewable infrastructure over the years
- Sizing new data centers (renewable and IT infrastructure)
- Other types of environmental impact

References

Koot, Martijn and Fons Wijnhoven. “Usage impact on data center electricity needs: A system dynamic forecasting model.” In: *Applied Energy* 291 (2021), p. 116798. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2021.116798.

Nafus, Dawn, Eve M. Schooler, and Karly Ann Burch. “Carbon-Responsive Computing: Changing the Nexus between Energy and Computing.” In: *Energies* 14.21 (2021). ISSN: 1996-1073. DOI: 10.3390/en14216917. URL: https://www.mdpi.com/1996-1073/14/21/6917.

Vasconcelos, Miguel Felipe Silva, Daniel Cordeiro, and Fanny Dufossé. "Indirect Network Impact on the Energy Consumption in Multi-clouds for Follow-the-renewables Approaches." In: *11th International Conference on Smart Cities and Green ICT Systems*. 2022.

LP Overview

Data center power consumption:

$$P_k^d \leq Pre_k^d + Pgrid_k^d + Pdch_k^d - Pch_k^d \tag{1}$$

where Pch_k^d is the power to charge the battery at each time of time slot k on DC^d and $Pdch_k^d$ is the power to discharge the battery, Pre_k^d is the solar power produced, and $Pgrid_k^d$ is the power used from the local grid.

Workload:

$$w_k^d \leq C^d \tag{2}$$

where w_k^d is number of cores needed during the kth time slot on DC^d, and C^d is the number of cores within DC^d .

LP Overview

Batteries level of energy (B_k^d):

$$B_k^d = B_{k-1}^d + Pch_{k-1}^d \times \eta_{ch} \times \Delta t - \frac{Pdch_{k-1}^d}{\eta_{dch}} \times \Delta t \tag{3}$$

where η_{ch} is efficiency of the charge process and η_{dch} is the efficiency of the discharge process.

Solar power production:

$$Pre_k^d = I_k^d \times Apv^d \times \eta_{pv} \tag{4}$$

where I_k^d is the solar irradiance, Apv^d the PV panel area, and η_{pv} is the efficiency of PV module

LP Overview

Objective function:

$$\text{minimize} \sum_{k=0}^{K-1} \sum_{d=1}^{D} (FPgrid_k^d + FPpv_k^d) + \sum_{d=1}^{D} FPbat^d \tag{5}$$