THE UNIVERSITY OF MELBOURNE

# Geographical Load Balancing for Sustainable Cloud Data Centers

**Adel Nadjaran Toosi**

*Cloud Computing and Distributed Systems (CLOUDS) Laboratory,*
*School of Computing and Information Systems*
*The University of Melbourne, Australia*

*Email: anadjaran@unimelb.edu.au*
*Homepage: https://adelnadjarantoosi.info/*

MELBOURNE CLOUDS LAB

# Biography

- Postdoctoral Research Fellow, University of Melbourne, 2015-
- PhD, Computer Science and Software Engineering, 2015
  - CLOUDS lab, Computing and Information Systems, University of Melbourne
  - Thesis: "*On the Economics of Infrastructure as a Service Cloud Providers: Pricing, Markets, and Profit Maximization*"
- Research Interests
  - Cloud Computing, Software-Defined Networking (SDN), Energy Efficiency and Green Computing, Soft Computing and Machine Learning
  - Focused on Resource Provisioning and Scheduling in Distributed Systems
- Publications
  - Over **24** publications, **14** Journal Articles (**9 A/A*,** 2010 ERA Ranking), **10** Conference papers
  - *h*-index: **15** and **1000+** citations (*Src: Google Scholar*)
- Current Research
  - Traffic engineering for energy efficient consolidation of virtual machines in SDN-enabled clouds

# Outline

- Introduction
  - High energy consumption in data centers
  - Renewable energy sources
  - Challenges of using renewable energy
- Geographical Load Balancing (GLB)
- Optimal offline algorithm and its intractability
- A GLB framework for web applications
- Performance evaluation
  - Experimental Setup
  - Results
- Summary and future directions

# Online Activities Every Second

**3.4 MILLION EMAILS**
ARE EXCHANGED[a]

**6,000 TWEETS**
ARE SENT ON TWITTER[b]

**100,000 SEARCHES**
ARE MADE ON GOOGLE[c]

**600 ITEMS**
ARE SOLD ON AMAZON[d]

**70,000 LIKES**
ARE GENERATED ON FACEBOOK[e]

**250 SONGS**
ARE DOWNLOADED VIA ITUNES[f]

a https://www.lifewire.com/how-many-emails-are-sent-every-day-1171210
b http://www.internetlivestats.com/twitter-statistics/
c http://www.statisticbrain.com/google-searches/
d https://www.inc.com/tom-popomaronis/amazon-just-eclipsed-records-selling-over-600-items-per-second.html
e https://www.brandwatch.com/blog/47-facebook-statistics-2016/
f http://www.billboard.com/biz/articles/news/1538108/itunes-crosses-25-billion-songs-sold-now-sells-21-million-songs-a-day

# Digital Transformation

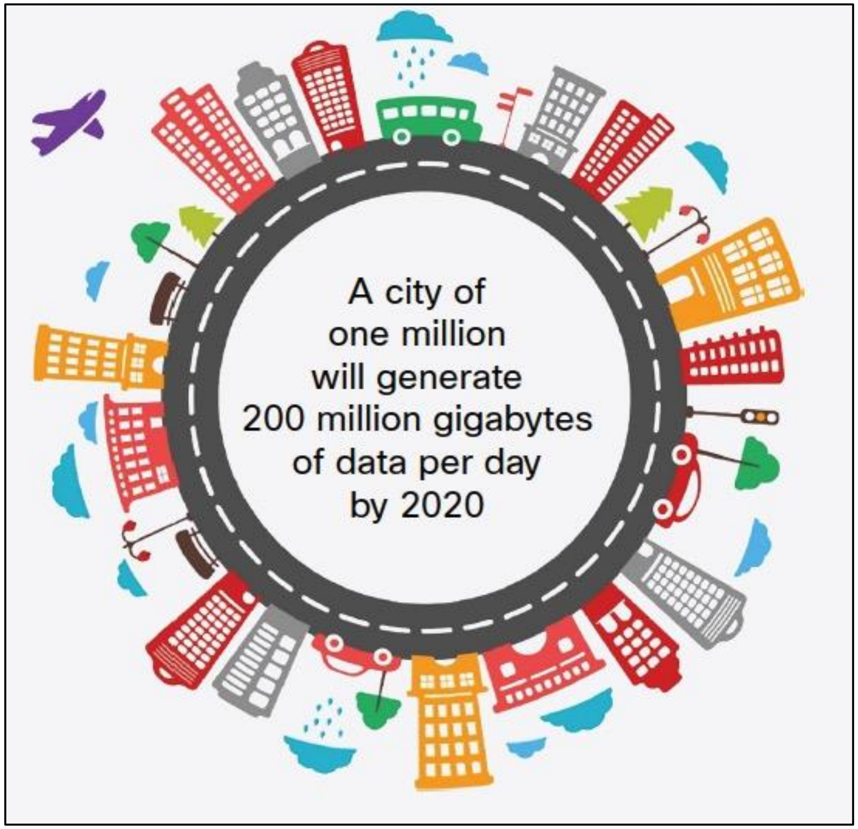
## More Cloud Apps
## More Data Centres

**5G and Smart Phones**
6B in 2020 (Ericsson)

**Smart Cities**

**Big data analytics Artificial Intelligence (AI)**

**Internet-of-Things (IoT) Devices**
25B in 2020 (Gartner)




Source: Cisco Global Cloud Index: Forecast and Methodology, 2015–2020

# Power Hungry Clouds

- Data centers used for hosting cloud applications consume large amounts of electricity
  - High **operational cost** for the cloud providers
  - High **carbon footprint** on the environment
- In 2014, US data centers alone consumed 70 billion kilowatt-hours of electricity
  - = Two-year power consumption of all households in **New York**
  - = The amount consumed by about **6.4 million** average American homes that year
  - This is projected to be responsible for the emission of nearly **50 million tons of carbon** pollution per annum in 2020.
    - *Source: US Natural Resources Defense Council (NRDC)*

# Renewable Energy

- Cloud providers aims
  - Reducing energy consumption
  - Dependence on brown energy
- Using renewable energy
  - On-site green power generation
  - Google, Microsoft and Amazon

*"Amazon Web Services (AWS) has built a wind farm in 2016 to exceed the goal of 40% electrical usage from renewable energy sources"*

Source: https://aws.amazon.com/about-aws/sustainability/

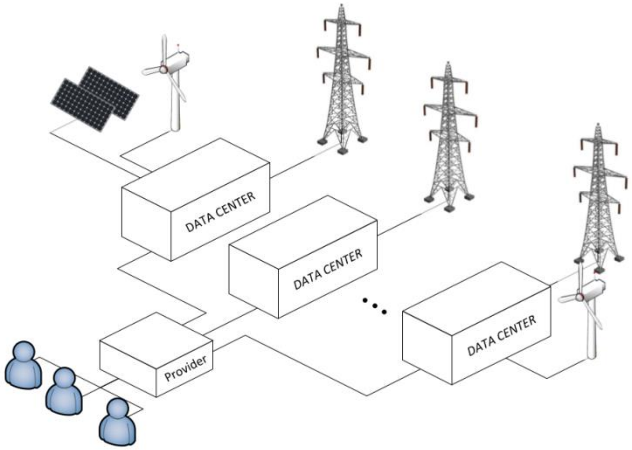
# Challenges

- ***Intermittency*** and ***unpredictability*** of renewable energy sources (Wind and Solar):
  - Powering data centers entirely with renewable energy sources is difficult
  - Mixed sources of energy for data centers:
    - » Grid power or brown energy
    - » Renewable energy sources or green energy
- Challenges:
  - Minimizing brown energy usage
  - Maximizing renewable energy utilization

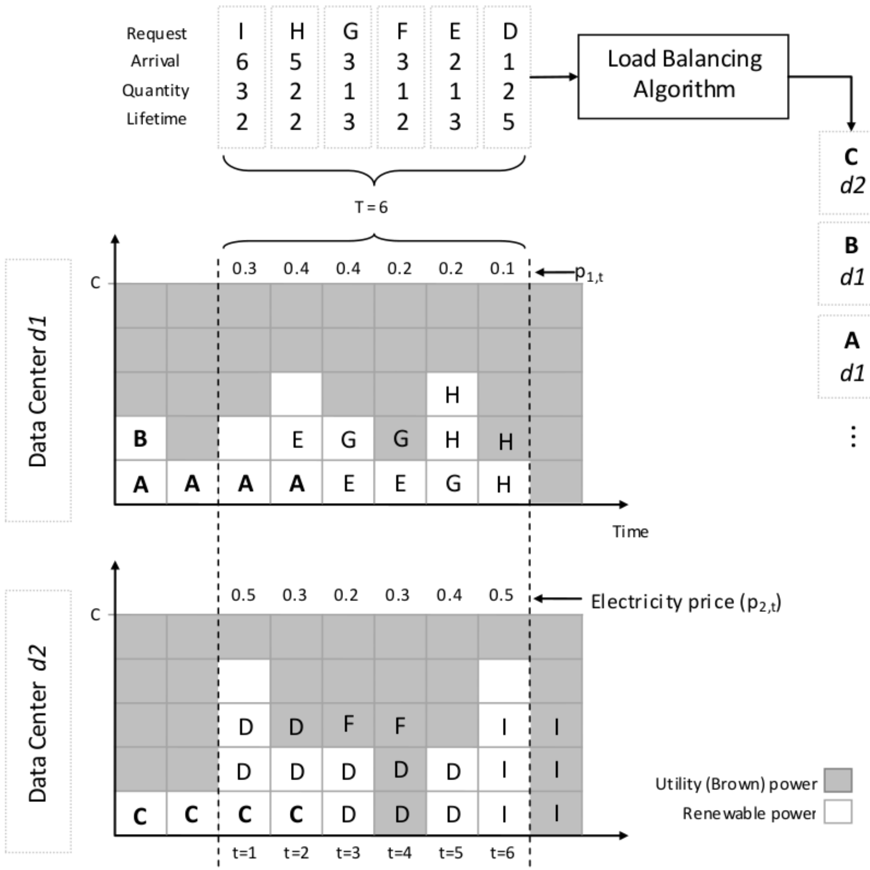
# Geographical Load Balancing (GLB)

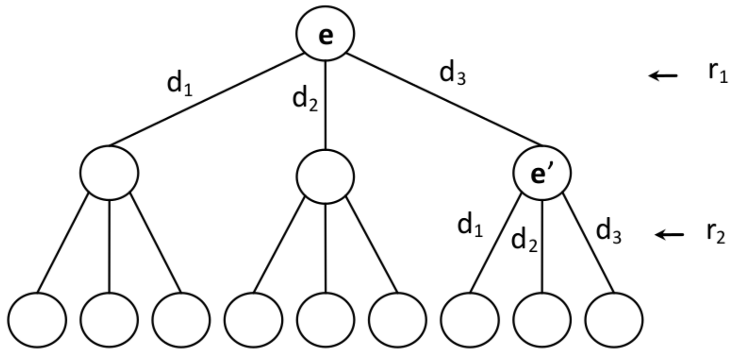
- Geographical load balancing (GLB) potentials:
    - **Follow-the-renewables**
- GLB approach benefits cloud providers but it raises an interesting, and challenging question:

*"With limited or even no a priori knowledge of the* ***future workload****, and the dynamic and unpredictable nature of renewable energy sources, how does one allocate requests to each data center such that the total* ***cost of power consumption*** *is minimized and the overall* ***renewable energy utilization*** *is maximized?"*



# Example: Offline GLB Problem



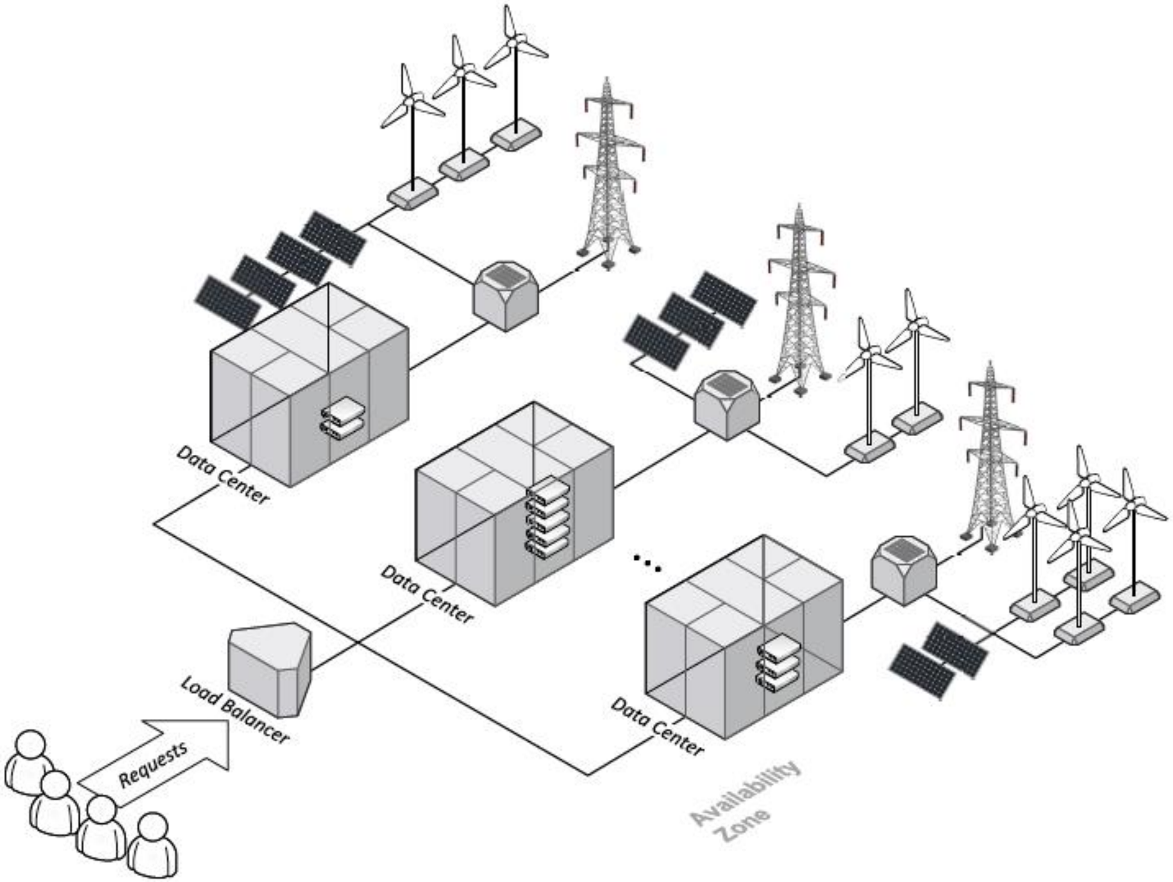



# Optimal Offline Algorithm

- Assuming the following information is known for a time window:
  - Future knowledge of renewable energy availability
  - Workload (i.e., number of requests, arrival time, and duration of requests)
- We showed that the optimal strategy is computationally intractable (Exponential time complexity)
  - Formal proof can be found in our paper.

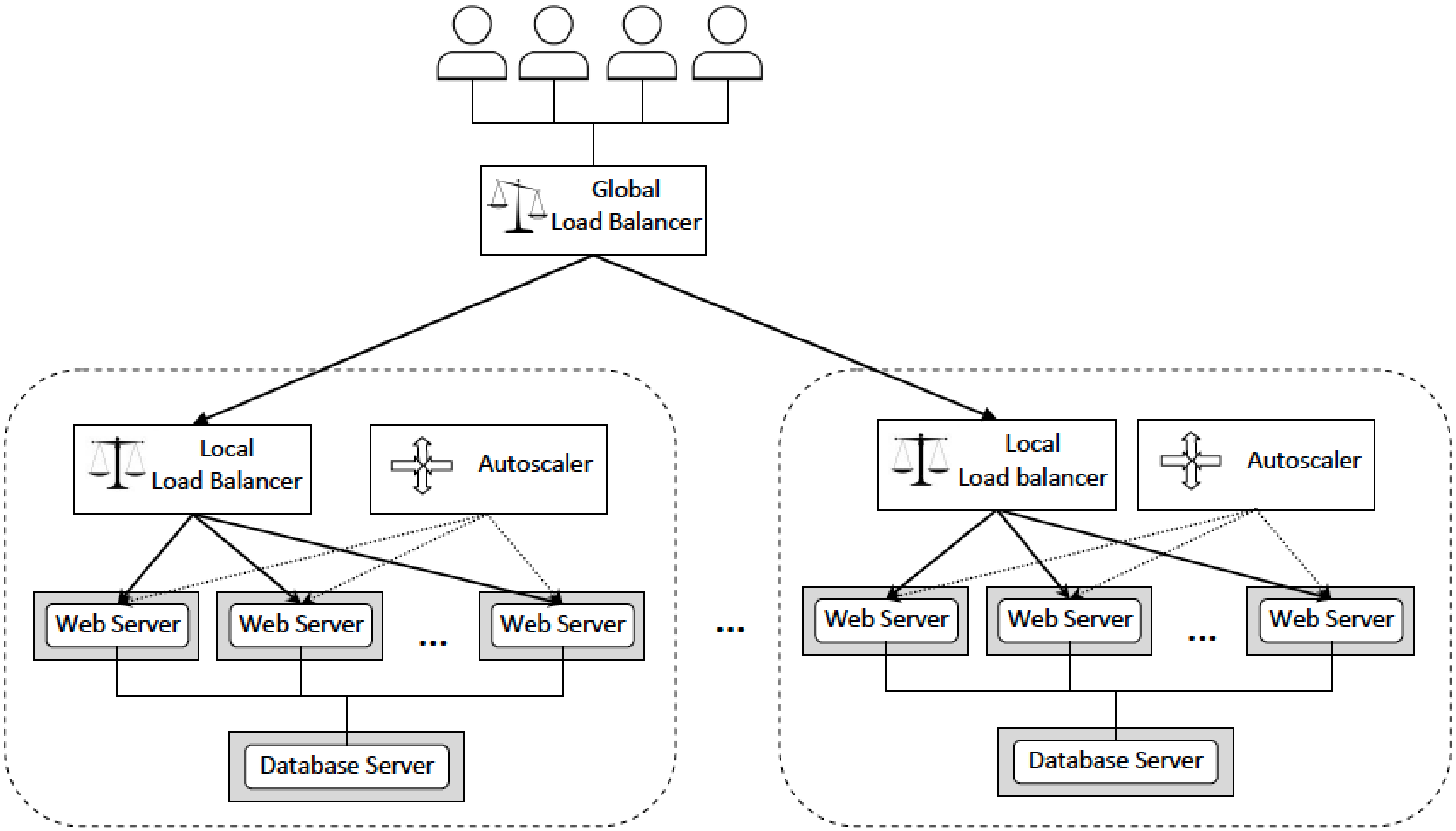
Adel Nadjaran Toosi and Rajkumar Buyya, A Fuzzy Logic-based Controller for Cost and Energy Efficient Load Balancing in Geo-Distributed Data Centers, In proceedings of the 8th IEEE/ACM International Conference on Utility and Cloud Computing (UCC'15), Limassol, Cyprus, Dec. 2015

# GLB for Web Applications

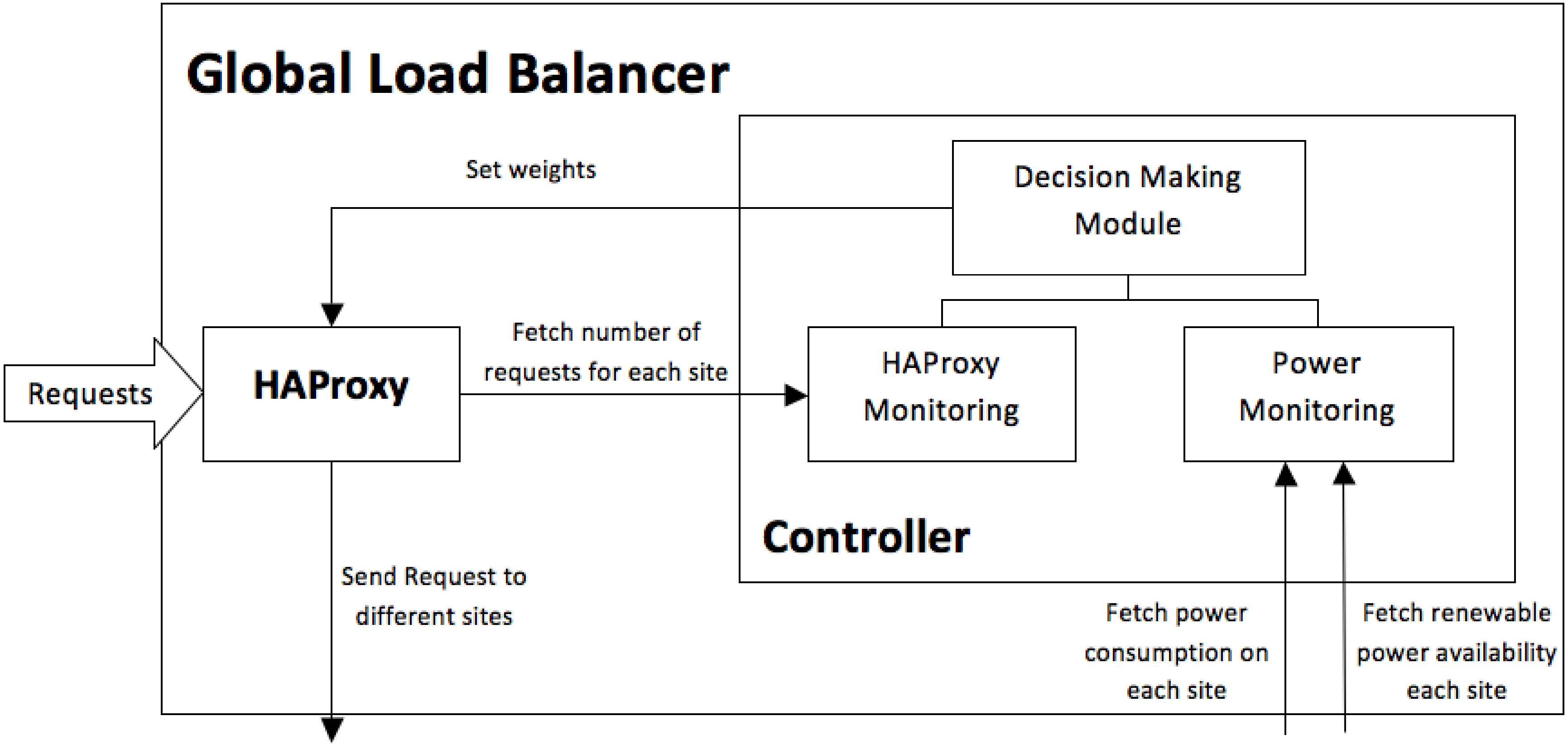



Adel Nadjaran Toosi, Chenhao Qu, Marcos Dias de Assuncao, and Rajkumar Buyya, Renewable-aware Geographical Load Balancing of Web Applications for Sustainable Data Centers, Journal of Network and Computer Applications (JNCA), Vol. 83, pp. 155-168, Apr. 2017.

# Overall System Architecture



# Global Load Balancer

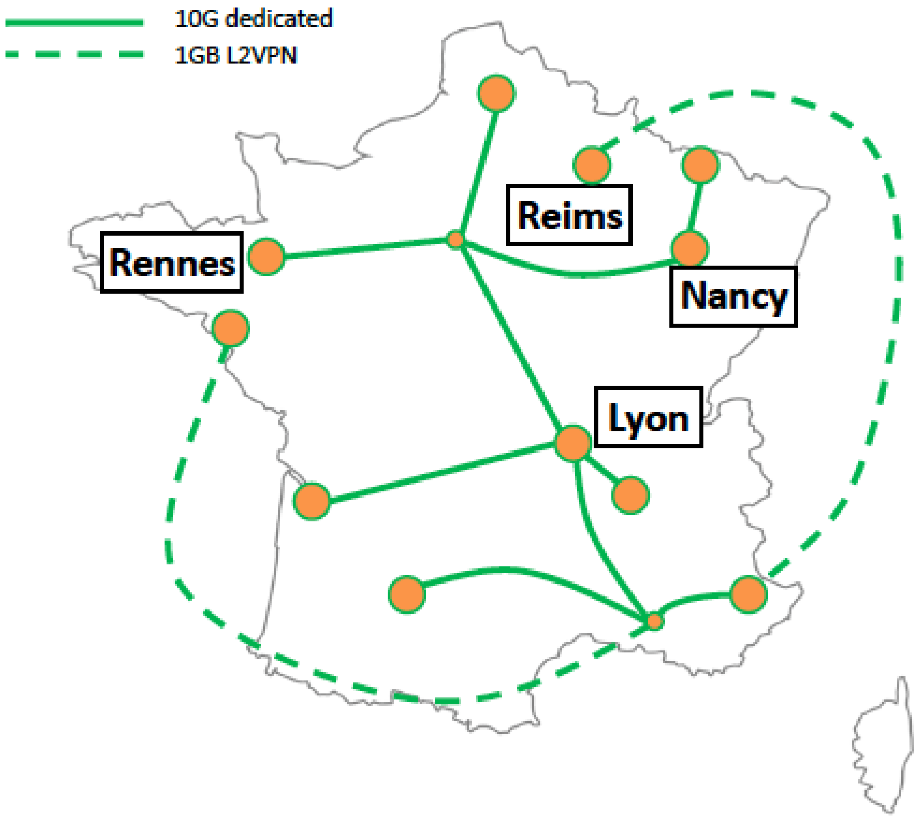


# Green Load Balancer (GreenLB)

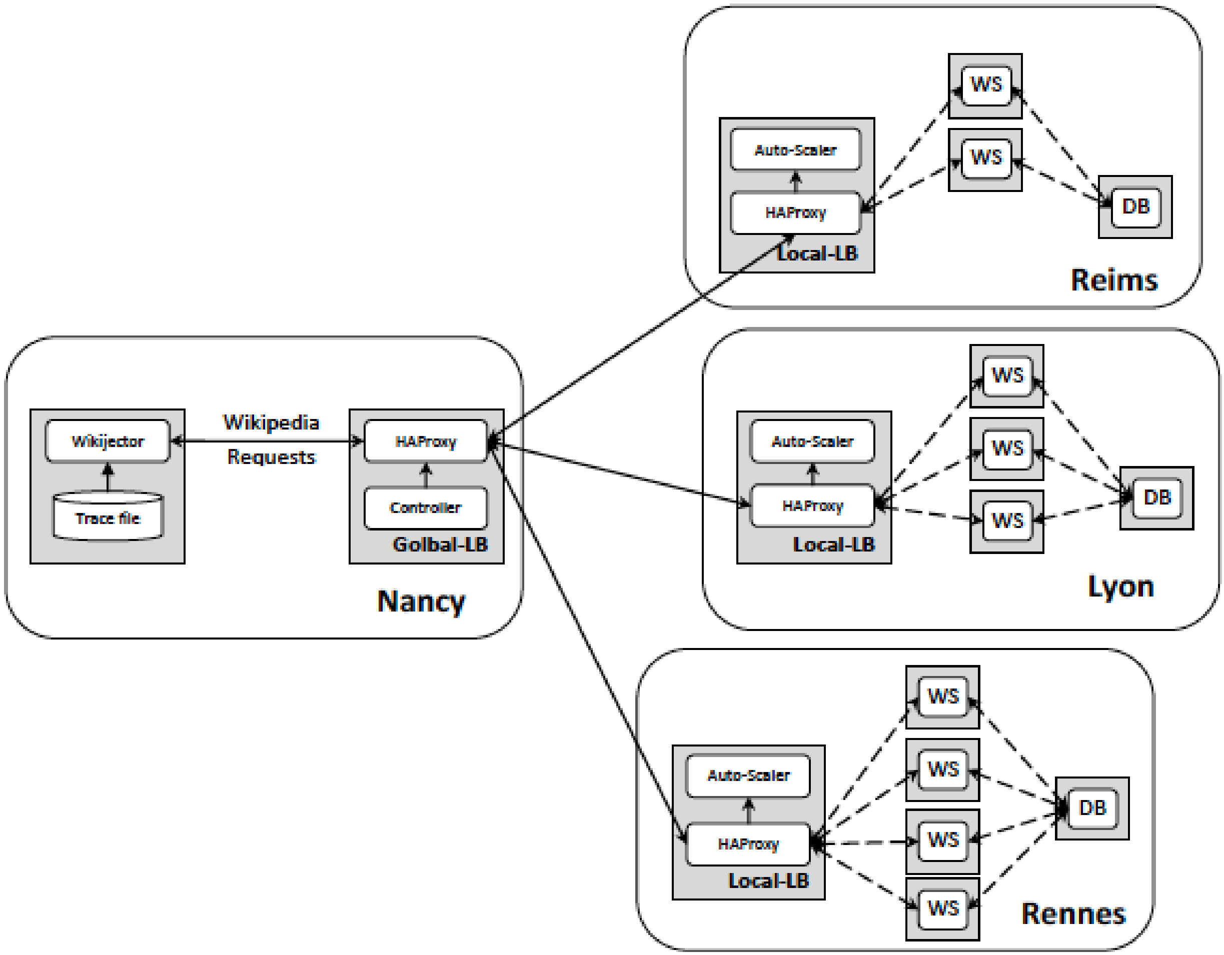
**Algorithm 1** Green Load Balancing (GreenLB) Policy

$R \leftarrow 0$
**for** all data centers $d$ in the list **do**
  $c \leftarrow$ Fetch the data center's energy consumption in Watt-hour within the time window
  $t \leftarrow$ Fetch the number of requests redirected to the site within the same time window
  $a \leftarrow$ Fetch currently available renewable power at the site in Watt
  $w \leftarrow$ Compute Watt-hour consumption per request $(c \div t)$
  $r_d \leftarrow$ Compute the request rate (#reqs/hour) data center $d$ can accommodate using renewables $(a \div w)$
  $R \leftarrow R + r_d$
**end for**
$\gamma \leftarrow$ Fetch request rate (#reqs/hour) at Global-LB
**if** $\gamma < R$ **then**
  **for** all data centers $d$ in the list **do**
    set weight as $r_d \div R$
  **end for**
**else**
  Find the data center $d'$ with the cheapest price of brown energy per request.
  $L \leftarrow \gamma$
  **for** all data centers $d$ in the list except $d'$ **do**
    set weight as $r_d \div \gamma$
    $L \leftarrow L - r$
  **end for**
  Set the weight for $d'$ as $L \div \gamma$
**end if**
Update HAProxy weights accordingly

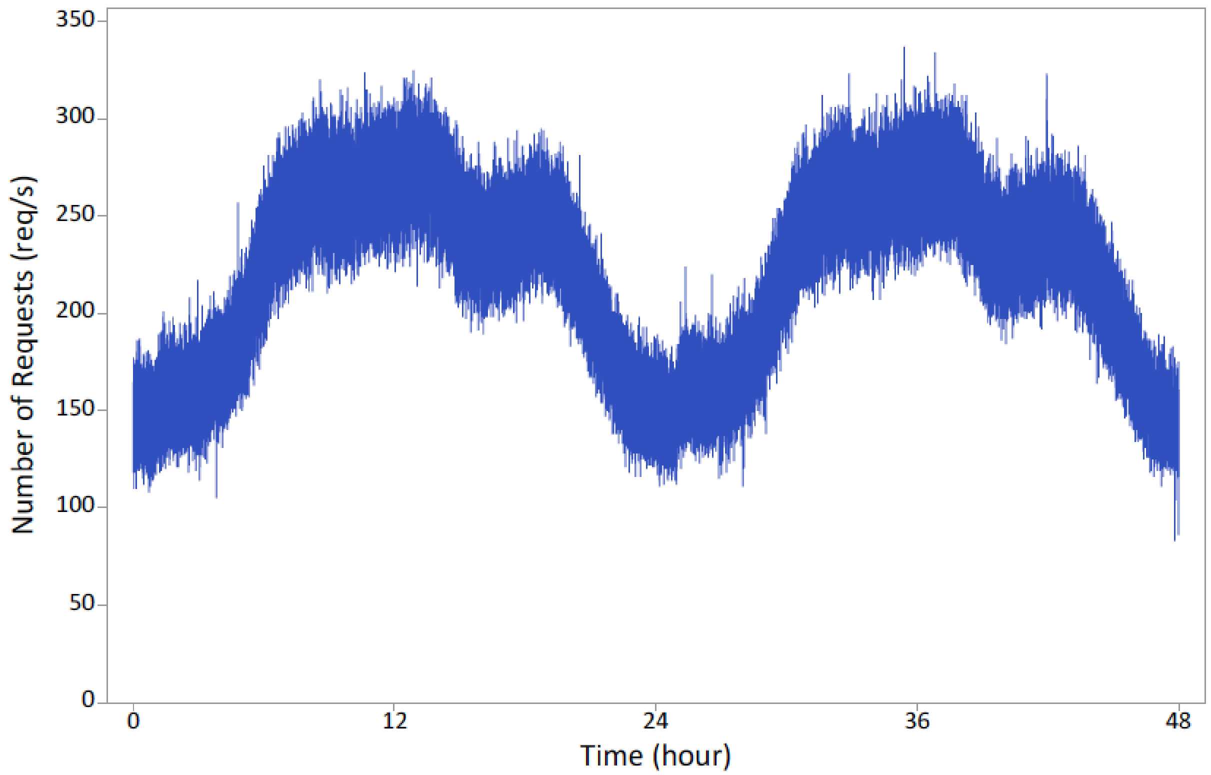
# Grid'5000 Testbed



# A Prototype System

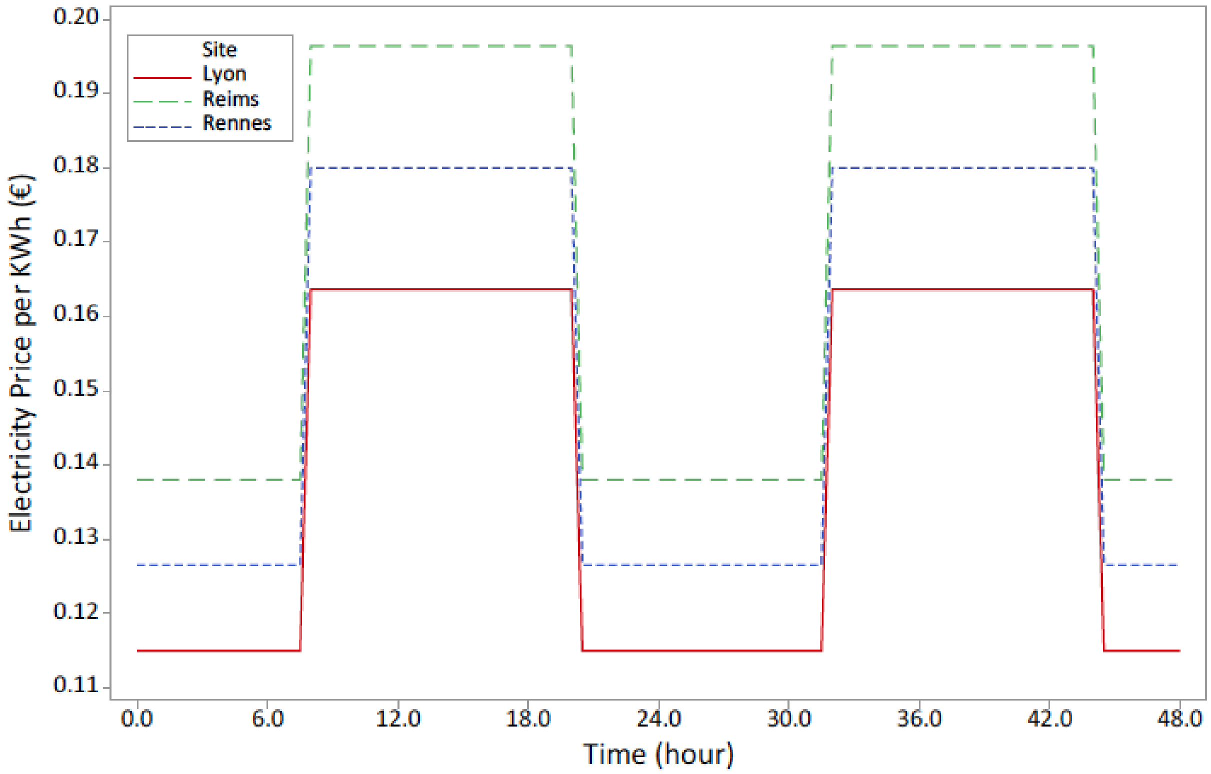


# Workload Traces

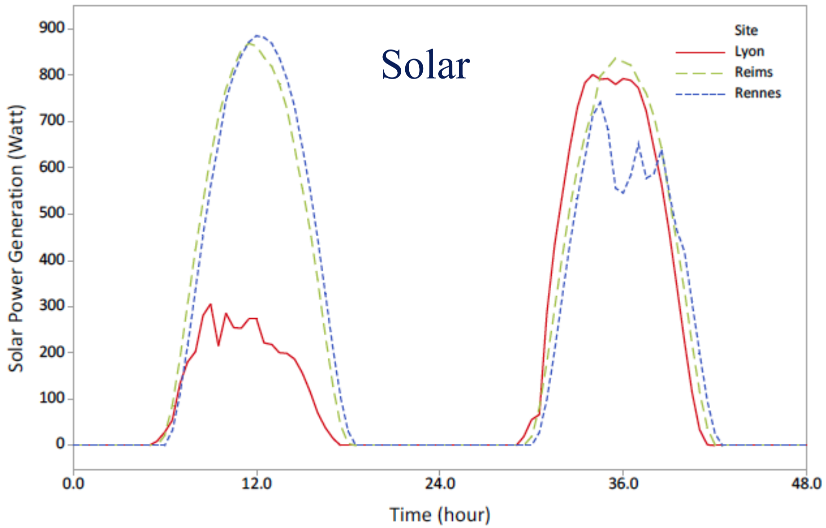


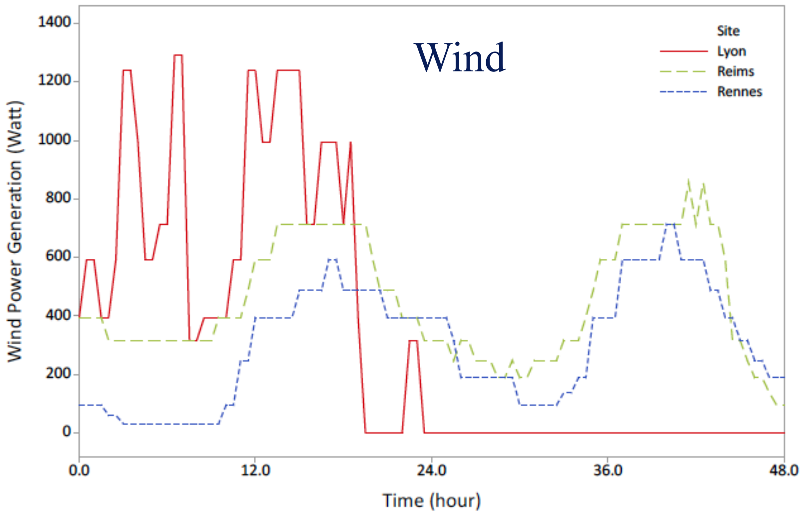

The 5% of English Wikipedia Requests for 19th and 20th of September 2007

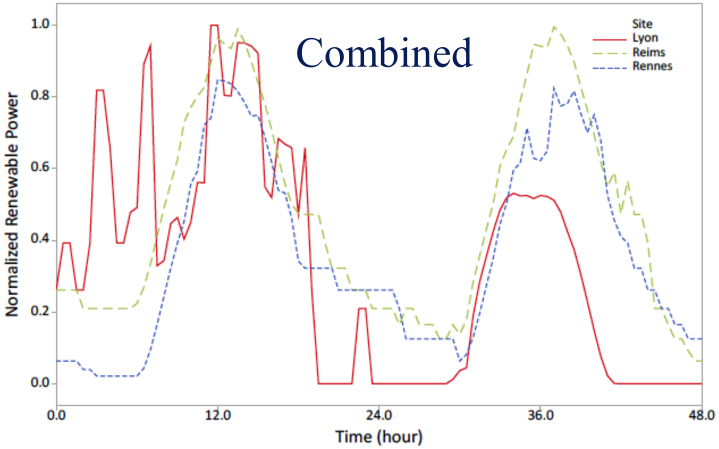
# Electricity Prices



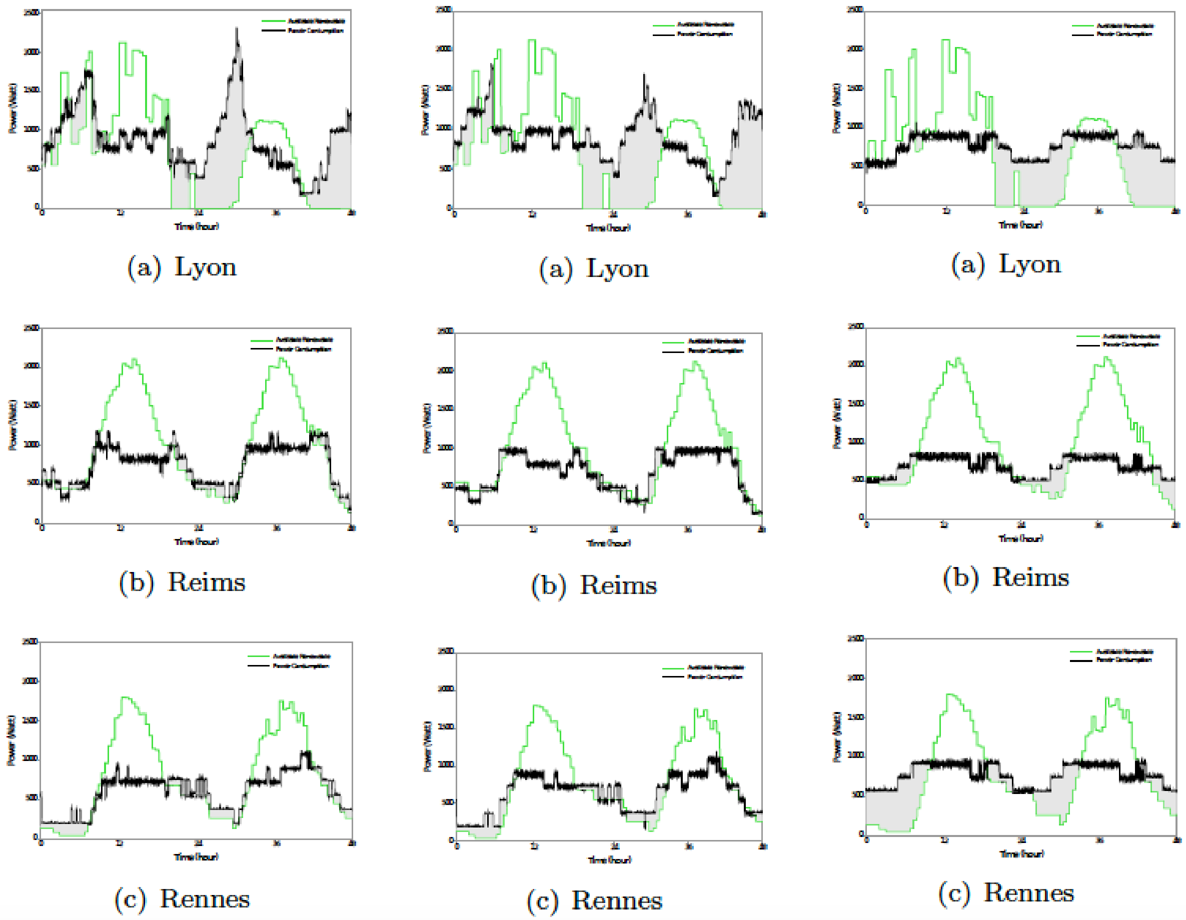
# Renewable Power Generation



# Results

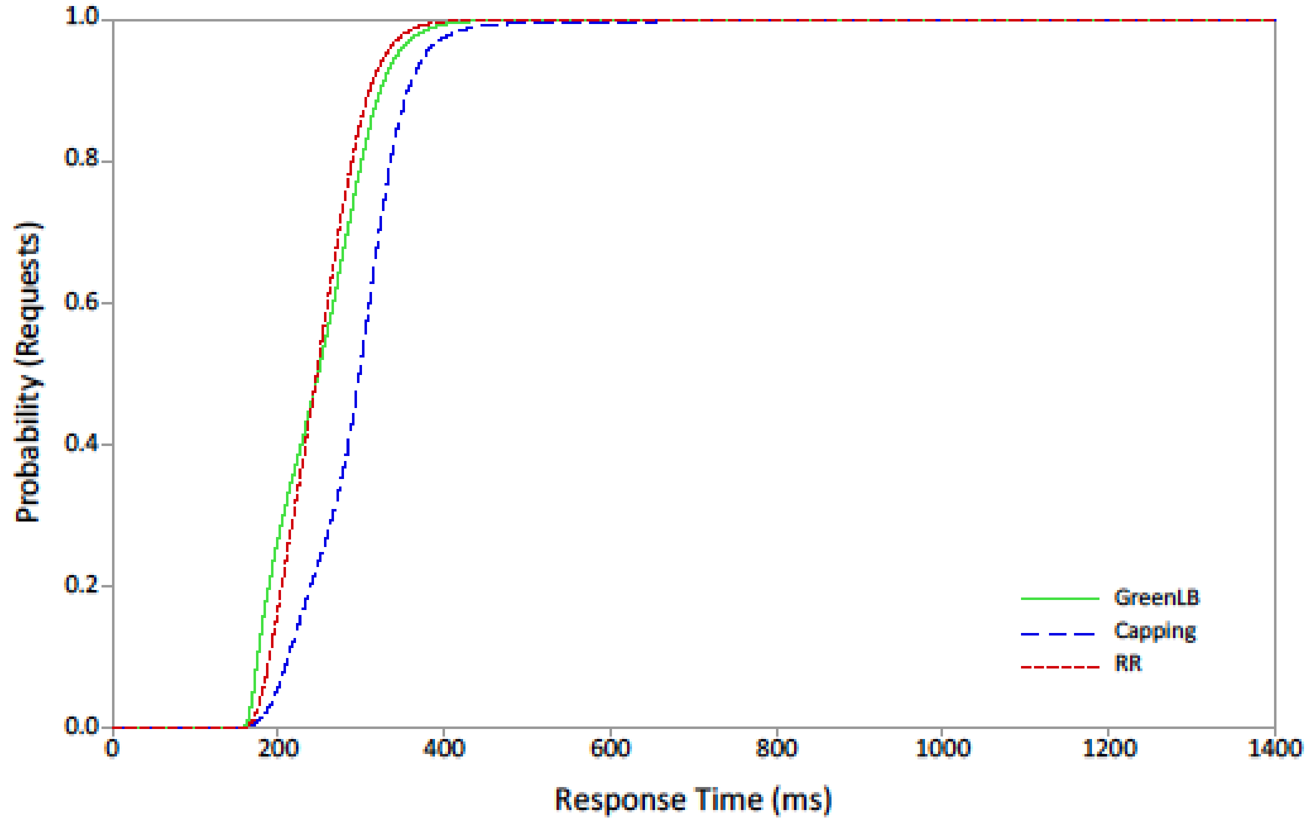



GreenLB | Capping | Round Robin

# Results

| Site | Metric | RR | Capping | GreenLB |
|---|---|---|---|---|
| **Lyon** | Power Consumption (kWh) | 36.3 | 42.9 | **41.2** |
| | Brown Consumption (kWh) | 13.3 | 19.0 | **16.9** |
| | Cost (€) | 1.71 | 2.31 | **2.01** |
| **Reims** | Power Consumption (kWh) | 32.5 | 32.5 | **35.4** |
| | Brown Consumption (kWh) | 2.1 | 1.1 | **1.9** |
| | Cost (€) | 0.42 | 0.15 | **0.27** |
| **Rennes** | Power Consumption (kWh) | 36.4 | 29.7 | **28.3** |
| | Brown Consumption (kWh) | 9.3 | 2.9 | **2.6** |
| | Cost (€) | 1.23 | 0.39 | **0.35** |
| **Total** | **Power Consumption (kWh)** | **105** | **105** | **105** |
| | **Brown Consumption (kWh)** | **25.7** | **23.0** | **21.4** |
| | **Cost (€)** | **3.36** | **2.85** | **2.63** |

# Results




CDF of average response time

# Summary

- A **cost** and **energy efficient** load balancing algorithm is proposed
    - Distributes web application requests among multiple data centres
- A prototype was implemented and experimental studies in a real testbed were performed
    - Real traces of web requests for English Wikipedia
    - Meteorological data in the location of each data centre to model solar and wind power generation
- It reduces **cost** by 22% and **brown energy** by 17% in comparison to a simple round robin policy
- It reduces **cost** by 8% and **brown energy** by 7% in comparison to a method by researchers from Rutgers and Princeton University

# Future Works

- Geographical load balancing for other types of workloads/applications
  - Bag of tasks, scientific workflows, map-reduce
- Demand response and capping the brown power consumption
  - To promote carbon neutrality
- "Sticky load balancing" policies
  - When a client and an application server connection is established, all subsequent requests from this session are redirected to the same server
  - Network proximity of the user

# THANK YOU

# Questions?