Energy-aware joint management of Networks and Cloud Infrastructures

IEEE Computer Networks 70 (2014) 75–95

GREEN CLOUD

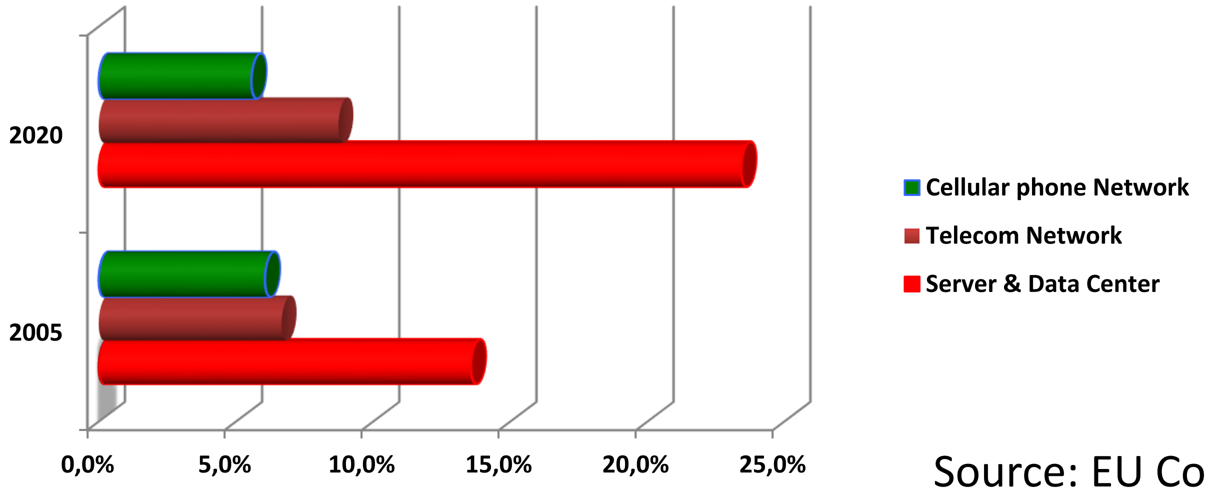
Bernardetta Addis
LORIA – INRIA Nancy Grand Est –
Universite de Lorraine
Giuliana Carello, Danilo Ardagna, Antonio Capone
DEI, Politecnico di Milano

Alta Scuola Politecnica Project Green Cloud

CLOUD CONSUMPTION

- About 0.5% of global electric **power consumption** is due to Data Centers (DC)
- In developed country:
  - UK: 2.2-3.3%
  - USA: 1.5%
- From the **environmental** point of view:
  - 2% of global CO2 emissions

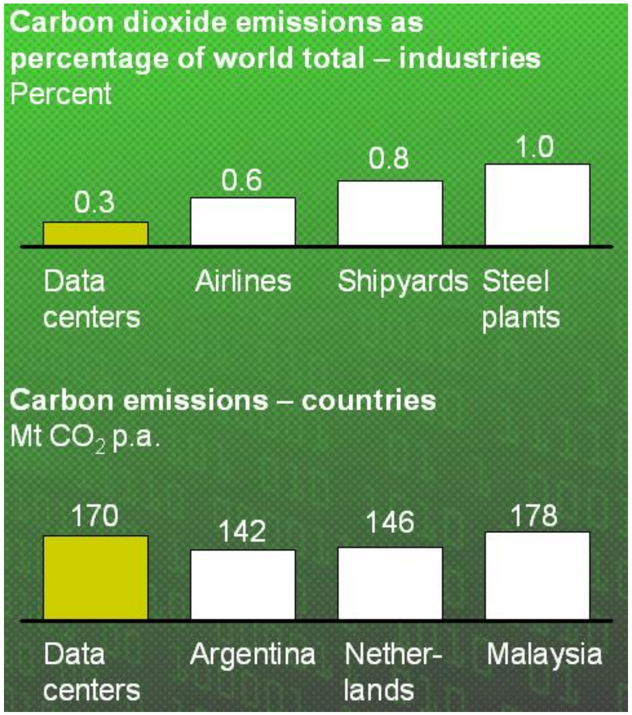
% European IT consumption



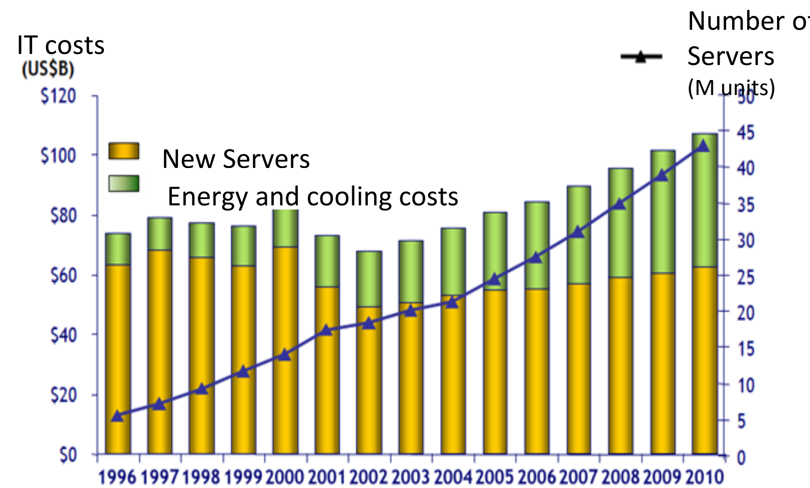
Source: EU Commission

ENVIRONMENTAL IMPACT

Environmental impact



...and costs



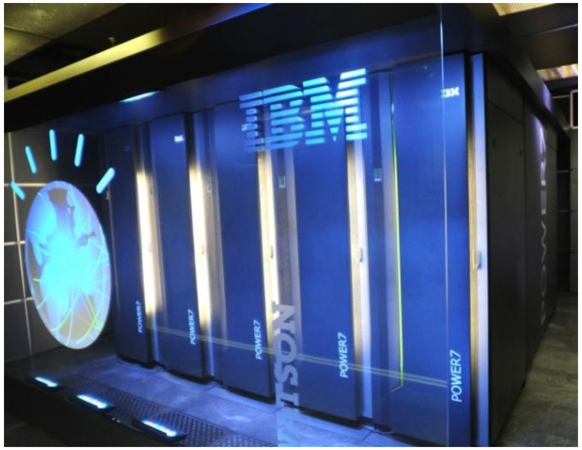
Source: EU Commission

- High consumption, environmental impact and energy costs of Cloud Computing

- A way out:
- **"GREEN" CLOUD COMPUTING**

GREEN CLOUD

OUR APPROACH



Green Cloud project:

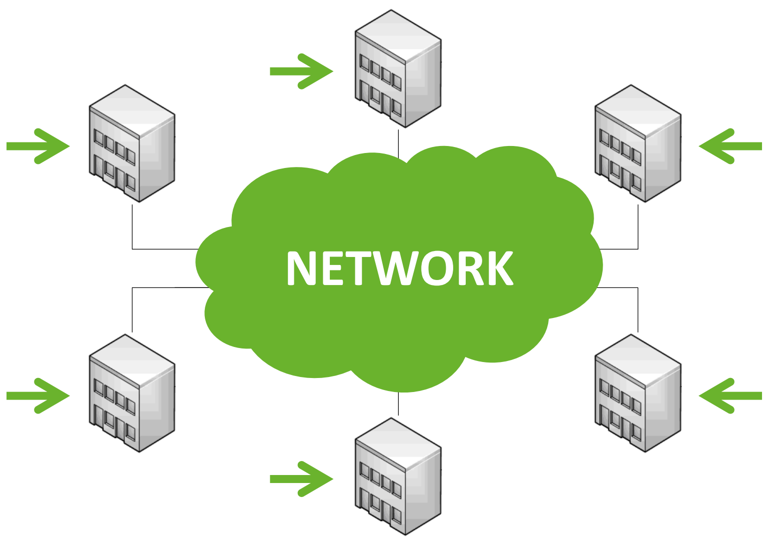
- Exploiting different energy costs and different workload due to **geographically distributed** Data Centers
- **Joint** Data Center and network management
- **Green energy sources** utilization

⇓

- Integer Linear Programming Optimization models
- Scenario definition and analysis

SCENARIO

- Optimization over **24 hours**
- Set of geographically distributed **DCs**
- **Traffic** profile for each Data Center over the day
- Possibility to **redirect** requests from one DC to another
- Fully connected **network**
- Energy **cost** for both DC and network



Request forwarding is due to:

- Lower energy cost
- Tradeoff between DC and network costs
- Available capacity

GIVEN

- The set of Data Centers: $\mathcal{N}$
- The set of request classes: $\mathcal{K}$
- The set of different type of servers: $\mathcal{L}$
- A planning horizon: $\mathcal{T}$

Each request

- is originated by a given DC
- can be served only by some type of servers

REDIRECTION VARIABLES:

$$x^t_{ijkl}$$

Requests rate (continuous variables) of the class *k* incoming to DC *i* served in DC *j* with a server of type *l*

CONSTRAINTS:

- REQUEST
  - The whole incoming traffic has to be served
  - Each request must be served by suitable server

$$\sum_{l \in \mathcal{L}} \sum_{j \in \mathcal{N}} x_{ijkl}^{t} = \lambda_{ik}^{t} \qquad \forall\, i \in \mathcal{N}, \forall k \in \mathcal{K}, \forall t \in \mathcal{T}$$

$$x_{ijkl}^{t} \leq \lambda_{ik}^{t} m_{kl} \qquad \forall i, j \in \mathcal{N}, \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, \forall t \in \mathcal{T}$$

CONSTRAINTS:

- REQUEST
- DATA CENTERS
  - DCs can handle a finite number of requests
  - Server have a utilization limit

$$w_{il}^t \geq \sum_{j \in \mathcal{N}} \sum_{k \in \mathcal{K}} \frac{D_k x_{jikl}^t}{\overline{U}} \qquad \forall i \in \mathcal{N}, \forall l \in \mathcal{L}, \forall t \in \mathcal{T}$$

$$\sum_{l \in \mathcal{L}} P_{il} w_{il}^t \leq C y_i \qquad \forall i \in \mathcal{N}$$

CONSTRAINTS:

- REQUEST
- DATA CENTERS
- NETWORKS
  - Links have limited bandwidth

$$\sum_{k \in \mathcal{K}} b_k \sum_{l \in \mathcal{L}} x_{ijkl}^t \leq Q_{ij} z_{ij}^t \qquad \forall i, j \in \mathcal{N}, \forall t \in \mathcal{T}$$

$$z_{ij}^t = z_{ji}^t \qquad \forall i, j \in \mathcal{N}, \forall t \in \mathcal{T}$$

CONSTRAINTS:

- REQUEST
- DATA CENTERS
- NETWORKS
- SWITCHING

$$\overline{w}_{il}^{t} \geq w_{il}^{t} - w_{il}^{t-1} \qquad \forall i \in \mathcal{N}, \forall l \in \mathcal{L}, \forall t \in \mathcal{T}$$

$$\underline{w}_{il}^{t} \geq w_{il}^{t-1} - w_{il}^{t} \qquad \forall i \in \mathcal{N}, \forall l \in \mathcal{L}, \forall t \in \mathcal{T}$$

$$\overline{z}_{il}^{t} \geq z_{il}^{t} - z_{il}^{t-1} \qquad \forall i \in \mathcal{N}, \forall l \in \mathcal{L}, \forall t \in \mathcal{T}$$

$$\underline{z}_{il}^{t} \geq z_{il}^{t-1} - z_{il}^{t} \qquad \forall i \in \mathcal{N}, \forall l \in \mathcal{L}, \forall t \in \mathcal{T}$$

BROWN MODEL

CONSTRAINTS:

- REQUEST
- DATA CENTERS
- NETWORKS
- SWITCHING

OBJECTIVE FUNCTION:

$$\min \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}}\left\{c_i^t\left[\rho_i\sum_{l\in\mathcal{L}}(\alpha_{il}w_{il}^t+\eta_{il}\overline{w}_{il}^t+\theta_{il}\underline{w}_{il}^t)\right]\right\}$$

$$+\sum_{t\in\mathcal{T}}\sum_{i,j\in\mathcal{N}}f_{ij}^tR_{ij}\left[\delta_{ij}z_{ij}^t+\tau_{ij}\overline{z}_{ij}^t+\xi_{ij}\underline{z}_{ij}^t+(\gamma_{ij}-\delta_{ij})\frac{\sum_{k\in\mathcal{K}}b_k\sum_{l\in\mathcal{L}}x_{ijkl}^t}{Q_{ij}}\right]$$

Minimization of the costs over 24 hours

- Cost for operating servers (w_{ij}^t)
- Cost for switching on/off servers $(\underline{w}_{ij}^t, \overline{w}_{ij}^t)$
- Bandwidth utilization
- Number of active link and link switching costs $(z_{ij}^t, \underline{z}_{ij}^t, \overline{z}_{ij}^t)$

ASSUMPTIONS:

- DCs can be powered through renewable sources:
  - Solar
  - Geothermic
  - Wind
- Limited amount of green energy available
- Only autonomous production
- Minor cost with respect to brown energy sources

ASSUMPTIONS:

- DCs can be powered through renewable sources:
  - Solar
  - Geothermic
  - Wind
- Limited amount of green energy available
- Only autonomous production
- Minor cost with respect to brown energy sources

IMPLICATIONS:

- Green powered DCs are preferred
- CO2 reduction

ASSUMPTIONS:

- DCs can be powered through renewable sources:
  - Solar
  - Geothermic
  - Wind
- Limited amount of green energy available
- Only autonomous production
- Minor cost with respect to brown energy sources

$$\min \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{N}}\left\{c_i^t\left[\rho_i\sum_{l\in\mathcal{L}}(\alpha_{il}w_{il}^t+\eta_{il}\overline{w}_{il}^t+\theta_{il}\underline{w}_{il}^t)-y_i^t\right]+g_i^t y_i^t\right\}$$

$$+\sum_{t\in\mathcal{T}}\sum_{i,j\in\mathcal{N}}f_{ij}^t R_{ij}\left[\delta_{ij}z_{ij}^t+\tau_{ij}\overline{z}_{ij}^t+\xi_{ij}\underline{z}_{ij}^t+(\gamma_{ij}-\delta_{ij})\frac{\sum_{k\in\mathcal{K}}b_k\sum_{l\in\mathcal{L}}x_{ijkl}^t}{Q_{ij}}\right]$$

IMPLICATIONS:

- Green powered DCs are preferred
- CO2 reduction

CONSTRAINTS:

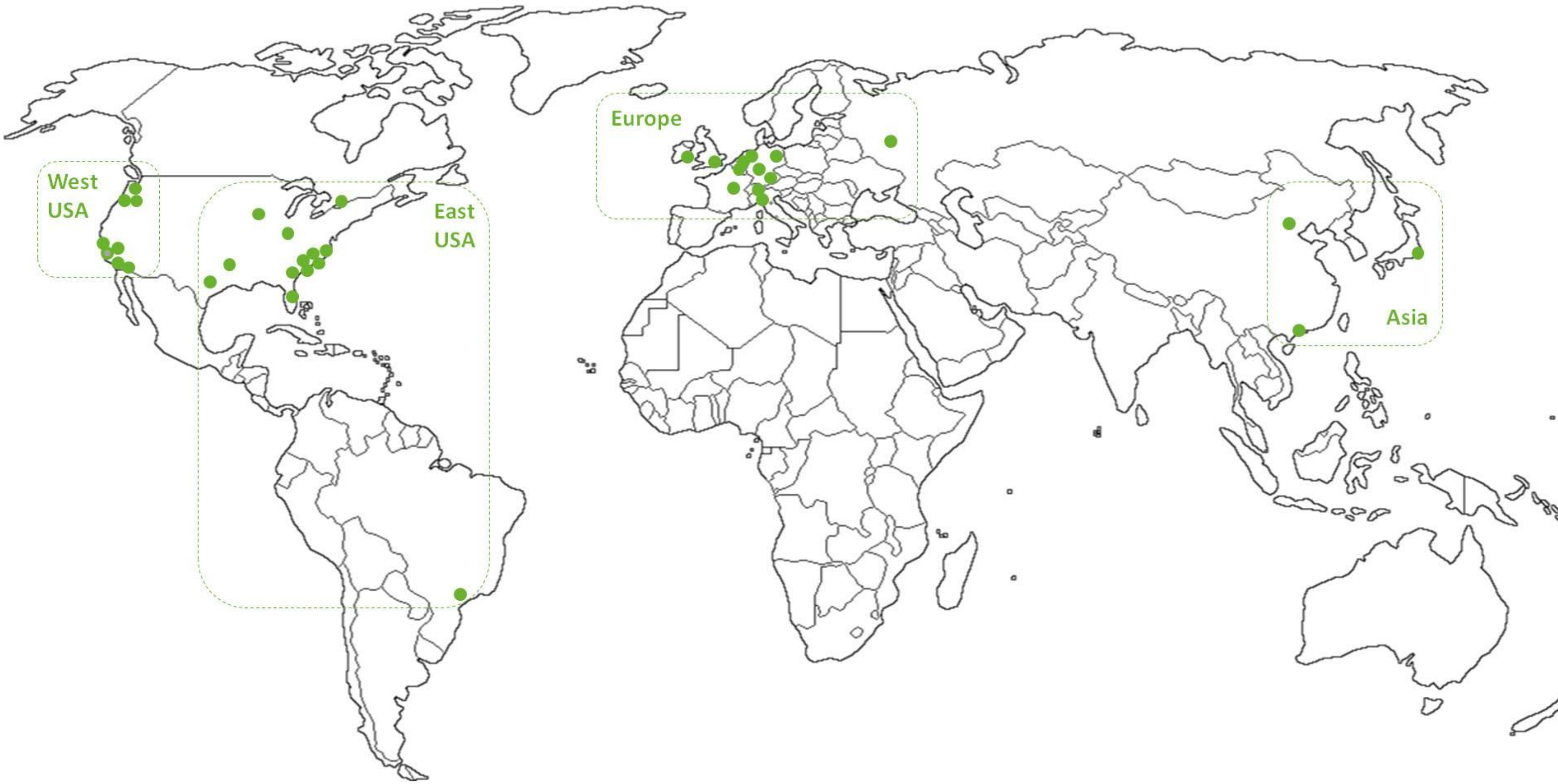
- REQUEST
- DATA CENTERS
- NETWORKS
- SWITCHING
- GREEN ENERGY

$$y_i^t \leq \rho_i \sum_{l \in \mathcal{L}} \left(\alpha_{il} w_{il}^t + \eta_{il} \overline{w}_{il}^t - \theta_{il} \underline{w}_{il}^t \right) \qquad \forall i \in \mathcal{N}, \forall t \in \mathcal{T}$$

$$y_i^t \leq \Gamma_i^t \qquad \forall i \in \mathcal{N}, \forall t \in \mathcal{T}$$

DATA CENTER MAP

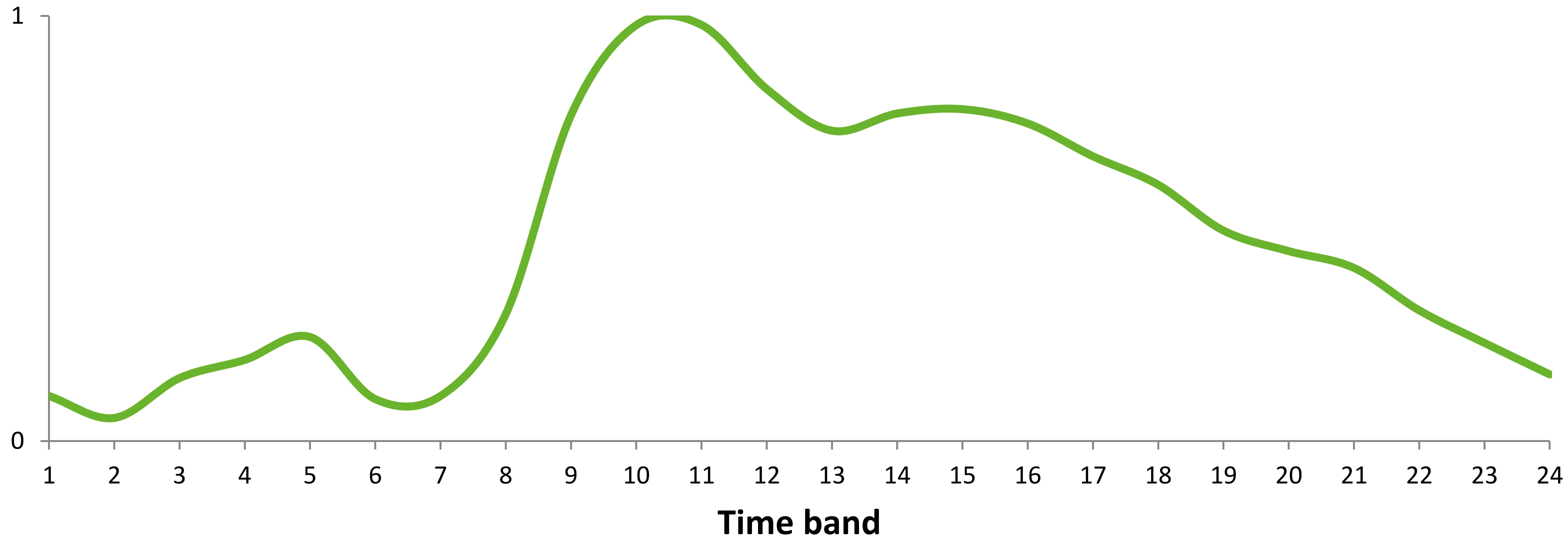
Google Data Center locations



INCOMING TRAFFIC

Starting from the **basic profile**:

- Re-scaling according to the geographical location
- Temporal shift to consider time zone differences



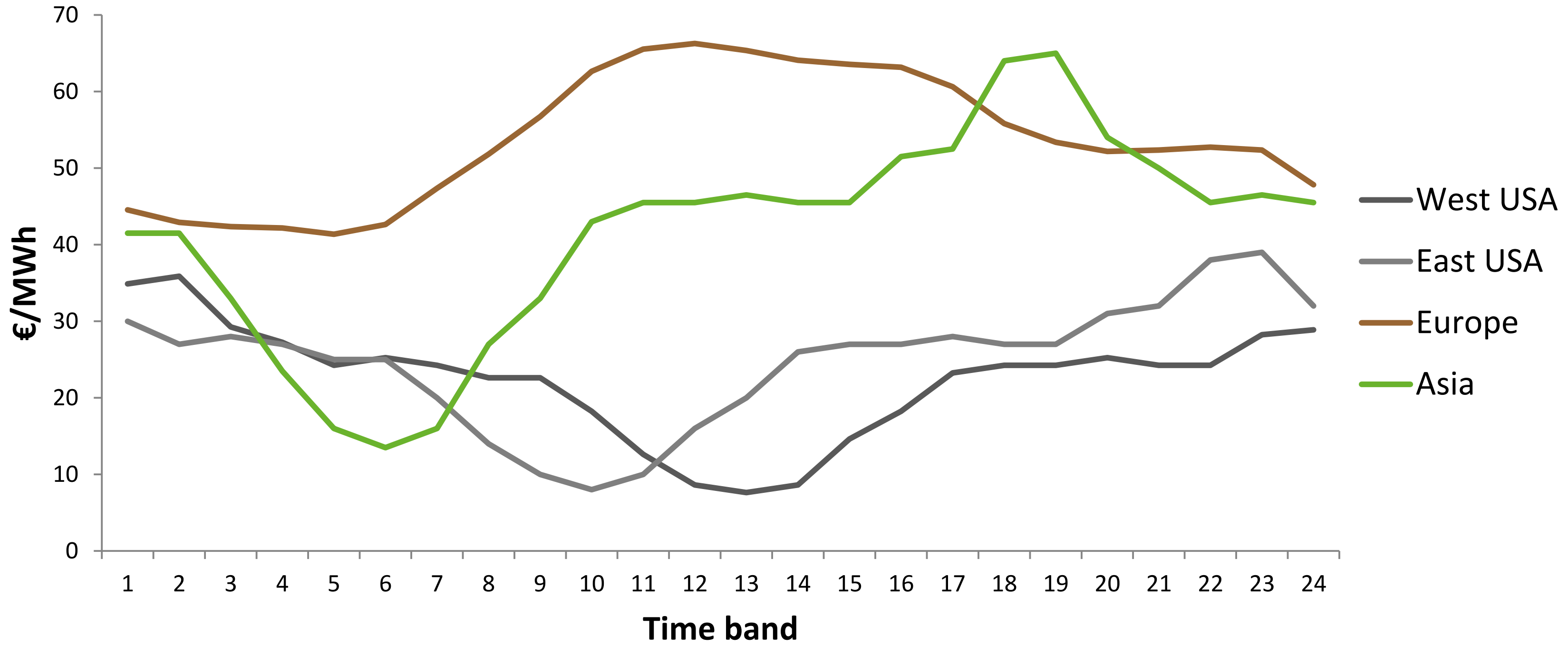
- Analysis of the model behavior with respect to a growing number of incoming requests:

  2 – 3 – 16 – 30 – 40 **billions of daily requests**

ENERGY COST

Differential **trend** according to:

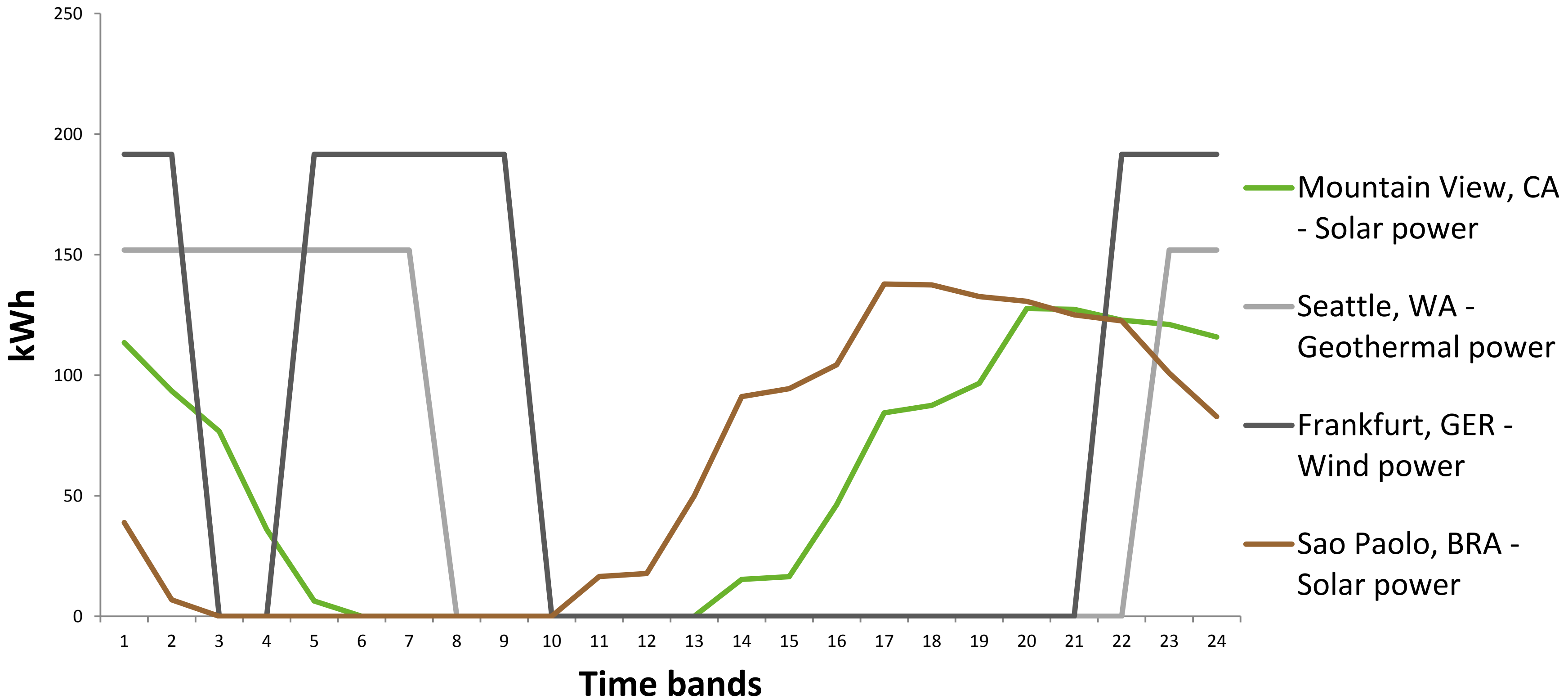
- Energy market studies
- Geographical location
- Time band



Location derived from **green energy sources** availability

GREEN ENERGY AVAILABILITY

- Different **trends** and **energy production** according to the different green energy sources
- Analysis based on the dimension and location of the DCs



POLICIES TO BE COMPARED

Base Case:

- Requests execution in local

Brown Model:

- Possibility of request redirection from one DC to another

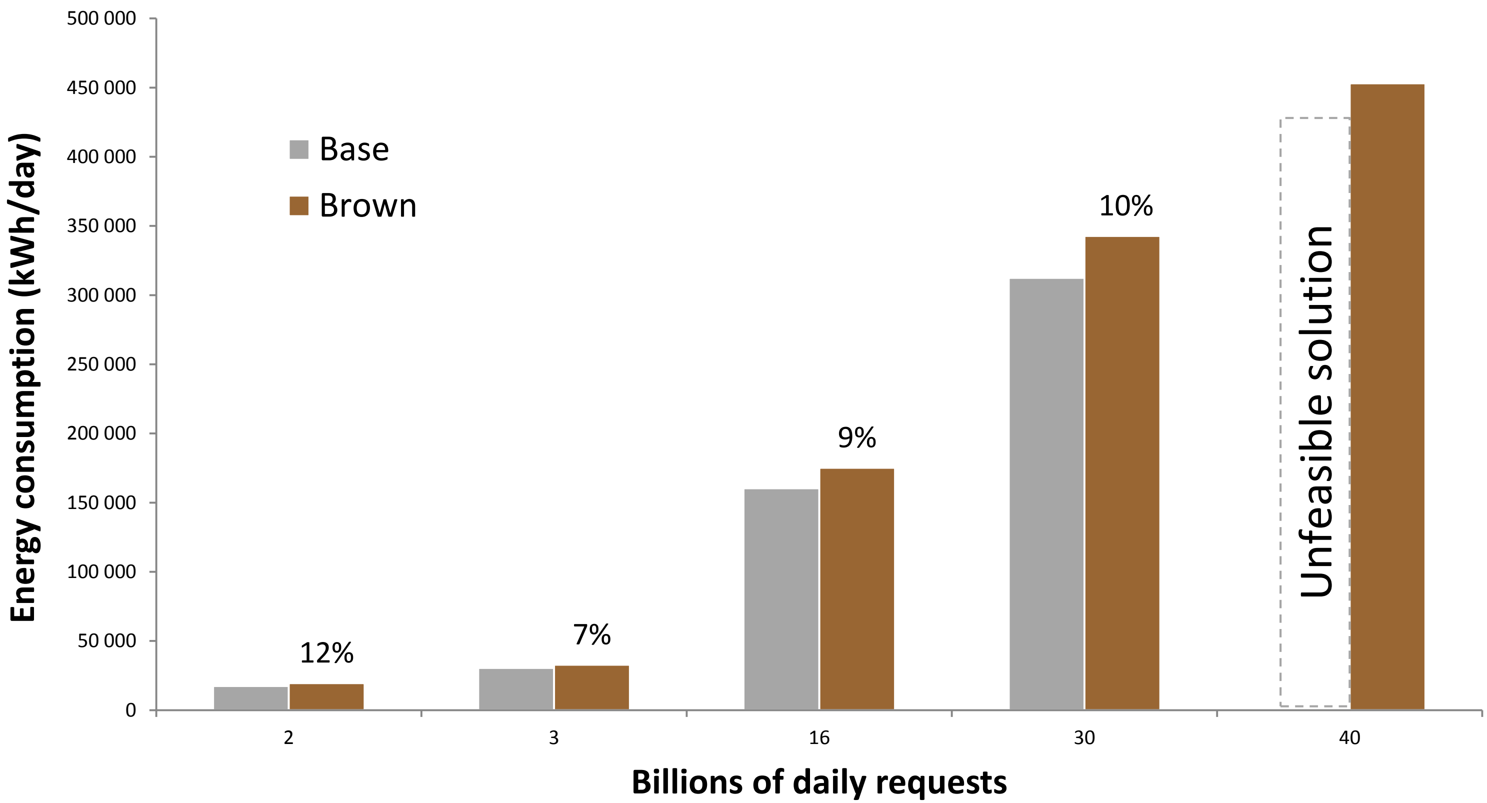
Green Model:

- Extension of the Brown Model taking into account green energy sources

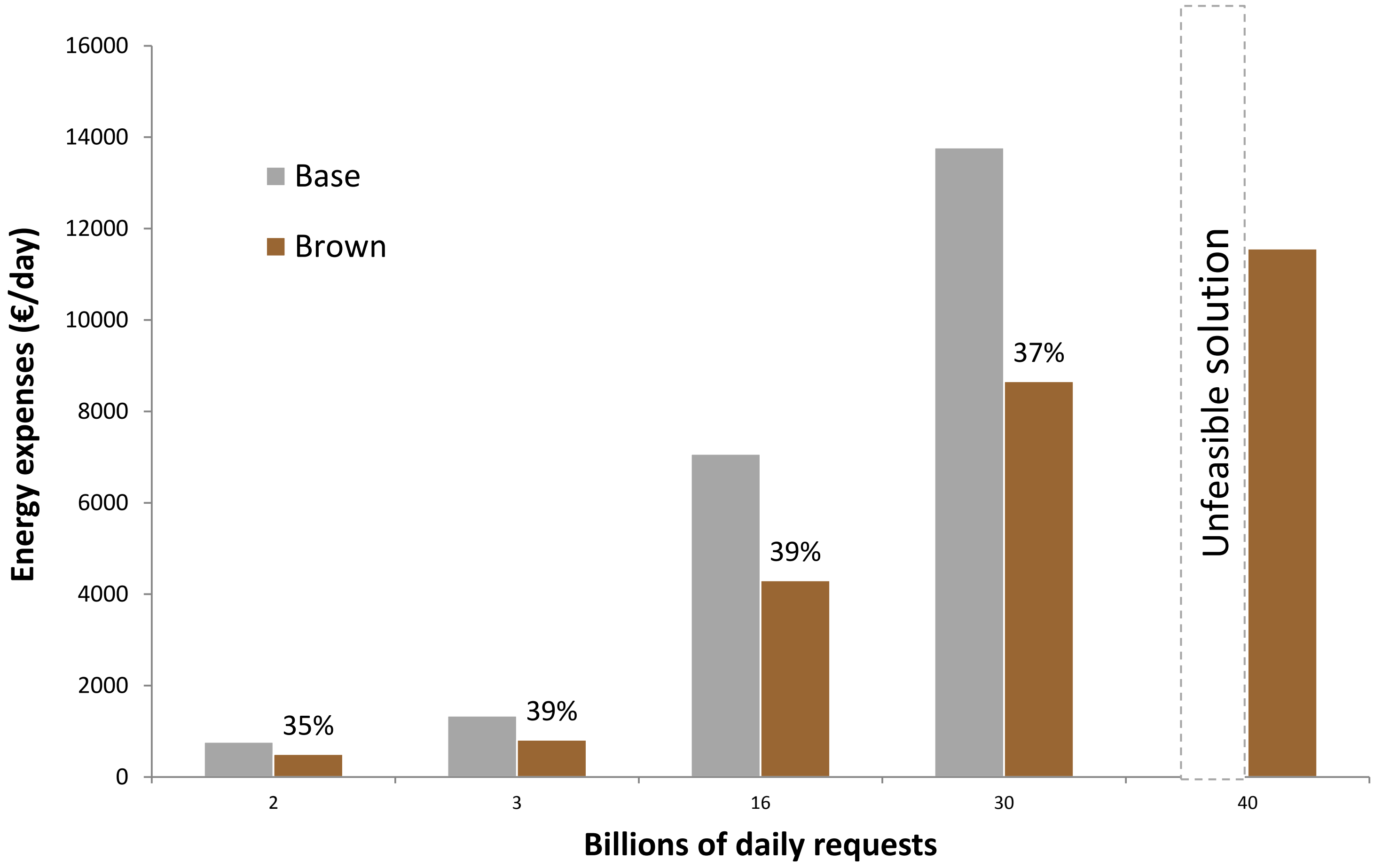
ENERGY CONSUMPTION

BROWN MODEL vs GREY

- Model priority: minimizing expenses
- Higher energy consumption due to network utilization for request forwarding

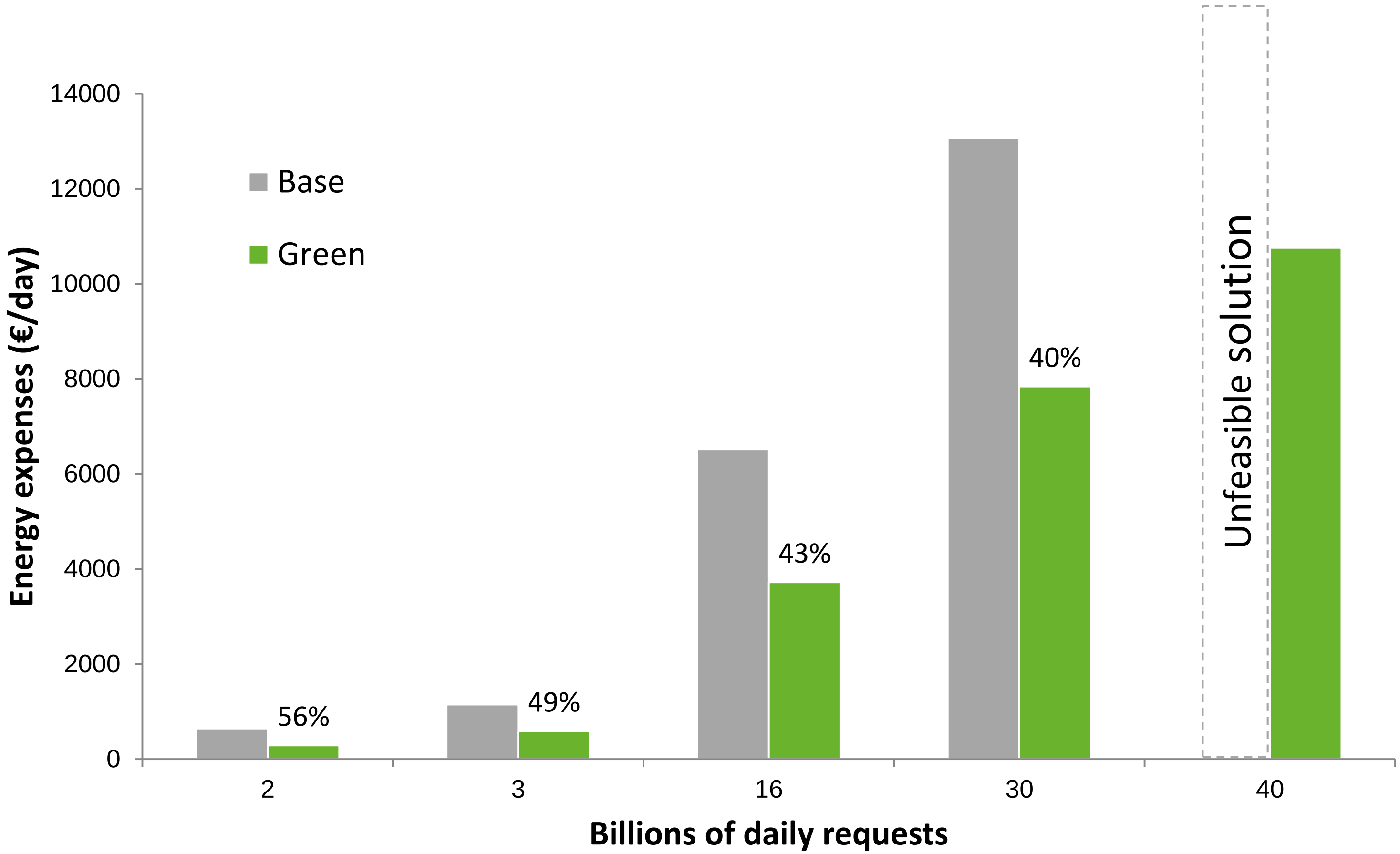


BROWN MODEL vs GREY



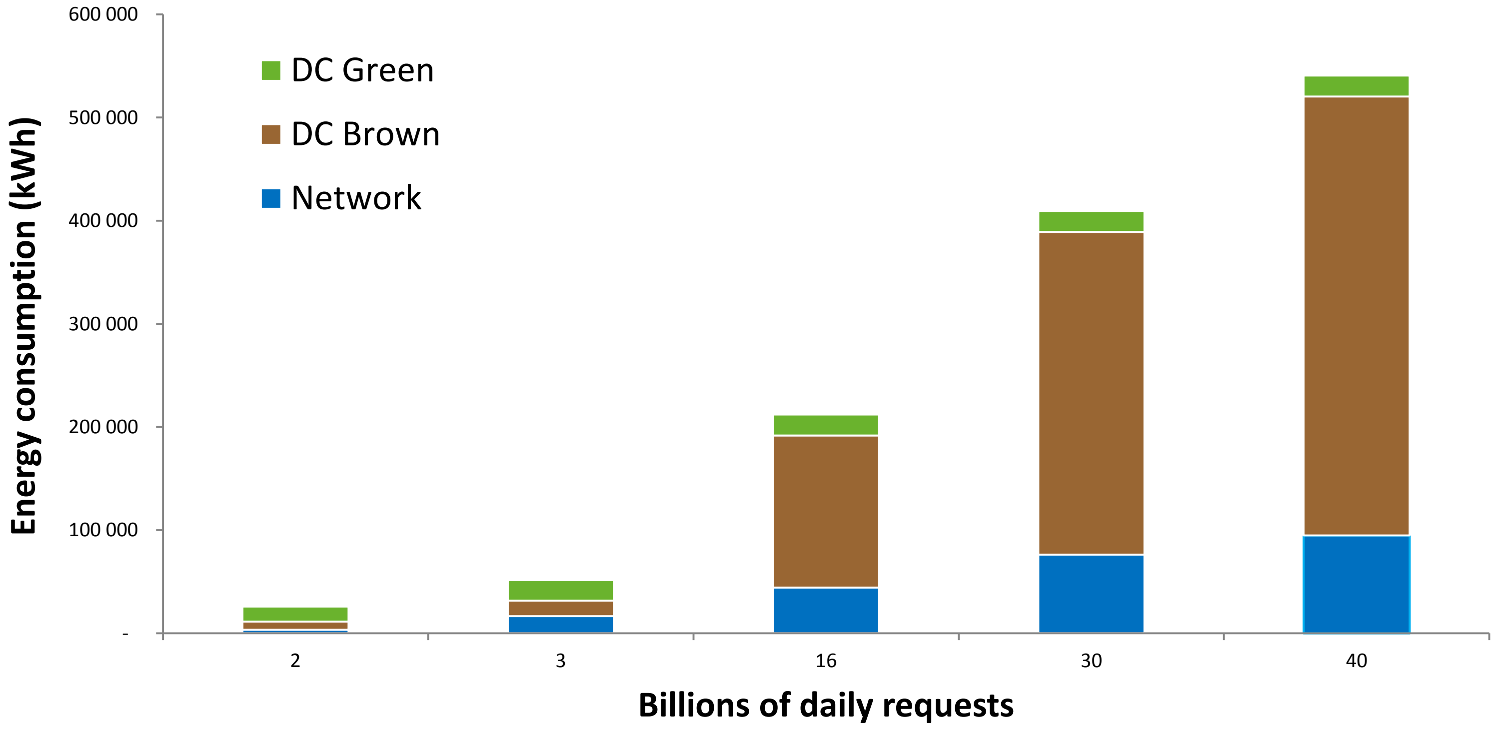
EXPENSES REDUCTION

GREEN MODEL vs GREY



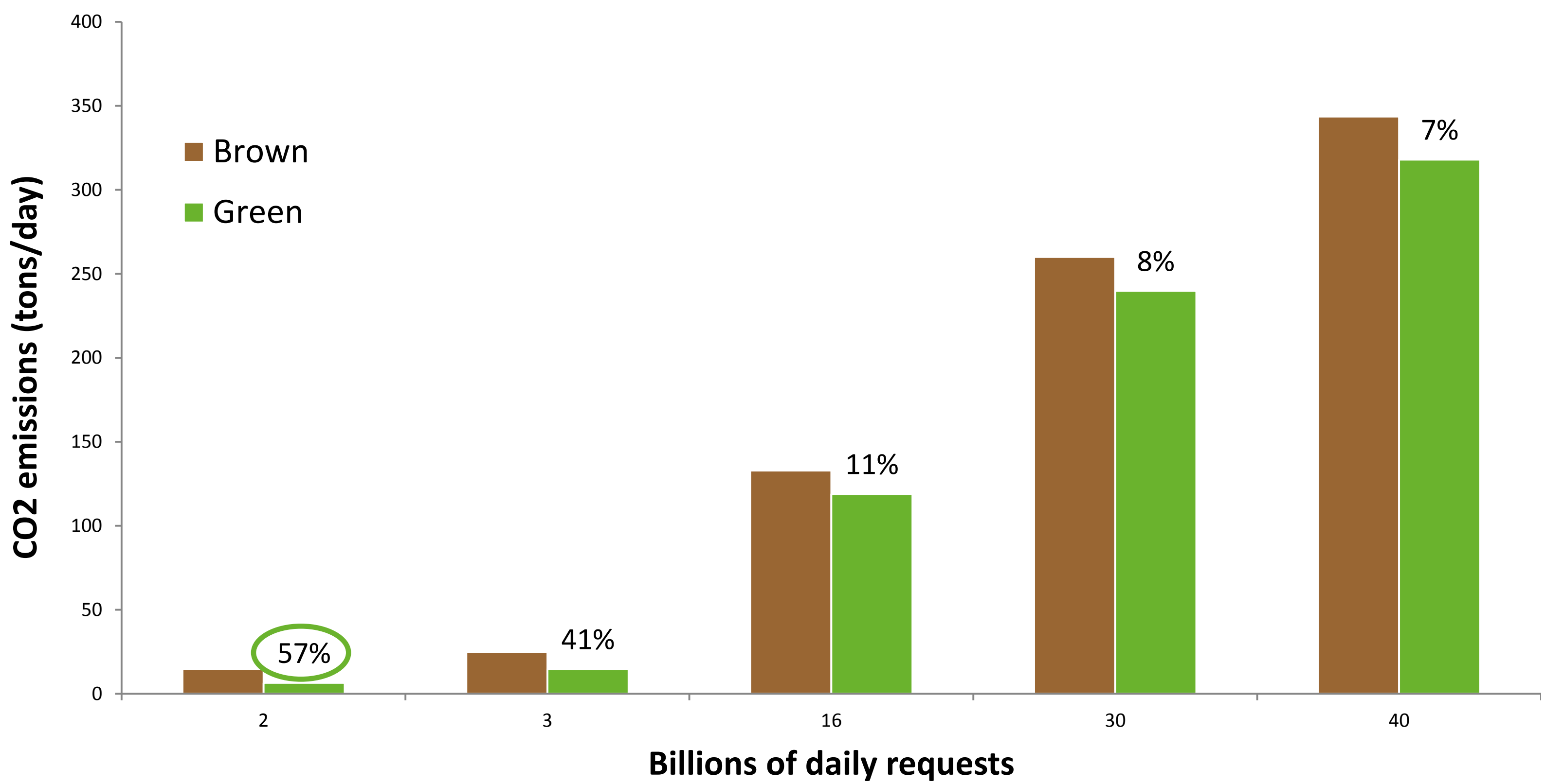
ENERGY CONSUMPTION REPARTITION

- Green energy **saturation**
- Possibility to better exploit the requests redirection by investing in green energy sources



CO_2 EMISSIONS

- Comparison of brown and green models based on the CO_2 emissions
- Maximum amount of green energy availability fixed at 8%



CONCLUSIONS

ECONOMICAL POINT OF VIEW:

- Optimization of Network and DCs jointly
- Expenses reduction up to 40%

ENVIRONMENTAL POINT OF VIEW:

- Reduction of greenhouse gas emissions
- Optimizing costs linked to the Carbon Credit system
- Promoting the use of green energy sources

FUTURE DEVELOPMENT

The **model** proved to be:

- Robust, flexible and scalable

POSSIBLE EXTENSIONS:

- Non fully connected network topology
- Service Level Agreement considerations and response time
- Tests on larger instances

GREEN CLOUD

THANK YOU

FOR THE ATTENTION!!!

ANY QUESTIONS?