Elastic operations in federated datacenters for performance and cost optimization

L. Velasco [a,*], A. Asensio [a], J.Ll. Berral [a,b], E. Bonetto [c], F. Musumeci [d], V. López [e]

[a] *Universitat Politècnica de Catalunya (UPC), Barcelona, Spain*
[b] *Barcelona Supercomputing Center (BSC), Barcelona, Spain*
[c] *Politecnico di Torino, Turin, Italy*
[d] *Politecnico di Milano, Milan, Italy*
[e] *Telefónica Investigación y Desarrollo (TID), Madrid, Spain*

Article history:
Available online 12 March 2014

1. Introduction

Cloud computing has transformed the IT industry, shaping the way IT hardware is designed and purchased [1]. Datacenters contain hardware and software to provide services over the Internet. Because datacenters consume huge amount of energy [2], energy expenditure becomes a predominant part of total operational expenditures for their operators. Aiming at reducing energy expenditure, datacenter operators can use, or even generate themselves, *green* energy coming from solar or wind sources; green energy would replace either partially or totally energy coming from *brown*, polluting sources. The drawback is that green energy is not always available, depending on the hour of the day, weather and season, among others. In contrast, brown energy can be drawn from the grid at any time, although its cost might vary along the day.

Thanks to virtualization, workloads (e.g. web applications) can be easily consolidated and placed in the most proper server according to its performance goals. By encapsulating workloads in virtual machines (VM) a datacenter resource manager can migrate them from one server to another looking for optimizing some objective function, such as energy consumption, whilst ensuring the committed quality of experience (QoE) [3,4].

Large Internet companies, such as Google and Microsoft, have their own infrastructures consisting in a number of large datacenters. These datacenters, placed in geographically diverse locations, guarantee good QoE to users and are interconnected through a wide area network [5]. Using that scheme, those companies can move workloads among datacenters to take advantage of reduced energy cost during off-peak energy periods in some locations (in addition to load balancing) while using green energy when is available in some other locations and turning off servers when they are not used, thus minimizing their energy expenditure.

* Corresponding author.
E-mail address: lvelasco@ac.upc.edu (L. Velasco).

Nonetheless, there is a large number of smaller independently operated infrastructures which cannot perform such elastic operations. Notwithstanding, those medium-size datacenters can cooperate by creating datacenter federations [6] to increase their revenue from using IT resources that would otherwise be underutilized, and to expand their geographic coverage without building new datacenters. Network providers can facilitate federated datacenters interconnection by allowing them to request connections' setup on demand with the desired bitrate, while tearing down those connections when they are not needed in a *pay as you go* model. To that end, network operators could use some automated interface to allow resource managers, in charge of each datacenter, to request such connections even in multi-domain network scenarios [7].

From the optical networking perspective, the advent of the flexgrid technology allows optical connections to be assigned an optical spectrum width according to their requested bitrate [8]. In addition, huge research and standardization work have been done defining control plane architectures and protocols to automate connection provisioning allowing to request them dynamically [9]. Led by the development of the software-defined network (SDN) concept, the IETF is also moving towards a centralized controller with the definition of the Application-Based Network Operations (ABNO) architecture [10]. In our previous works in [11,12], we studied the relation between datacenter management and flexgrid networks using the ABNO architecture.

In this work we assume a set of federated datacenters strategically placed around the globe so as to provide worldwide, high QoE services, interconnected by a flexgrid-based network. Each datacenter has access to some amount of energy coming from green sources which can cover some percentage of total energy consumption (*green coverage*), being the rest drawn from the grid. We study two approaches to orchestrate such datacenter federation to provide committed QoE while minimizing operational expenditures: distributed and centralized. In the distributed approach, datacenters schedule VM placement so as to minimize an estimation of the energy cost plus communication costs while ensuring QoE. In the centralized approach, a centralized orchestrator computes the global optima from placing VM to take full advantage from green energy availability in the federated datacenters.

The internal datacenters architecture has become crucial to deploy energy-efficient infrastructures. A certain number of switches is necessary to provide connectivity between servers in the datacenter and to interface the datacenter with the Internet. Consequently, according to the datacenter architecture being adopted, a corresponding power is consumed, basically dependent on the number and type of switches used. Among the various intra-datacenter architectures studied in literature (see [13] for a detailed survey), the so-called *flattened butterfly* architecture has been identified as the most power-efficient datacenter architecture, thanks to its power-proportional behavior, i.e. its power consumption is proportional to the number of currently used servers. However, the most widely-deployed architecture for datacenter is the so-called *fat-tree* topology [14], which is based on a hierarchical structure where large higher-order switches represent the interface of the datacenter towards the network infrastructure, and are connected to the servers via a series of lower-order switches, providing the intra-datacenter connectivity.

Since minimizing energy expenditures is really important for datacenter operators, many papers can be found in the literature partially addressing that problem [15–18]. In [15], the authors propose scheduling workload in a datacenter coinciding with the availability of green energy, consolidating all the jobs on time slots with solar energy available, increasing green energy consumption up to 31%. Authors in [16] present a datacenter architecture to reduce power consumption, while guarantee QoE. They consider online-monitoring and VM placement optimization achieving energy savings up to 27%. Other works, e.g. [17], refer to the problem of load balance datacenter workloads geographically, following green energy availability, to reduce the amount of brown energy consumed focusing mainly on wind energy and the capability of store energy. Other works focus on the importance of counting as "energy expenditure" every element in the datacenter, not only computing machinery. The author in [18] remarks the idea that all IT equipment counts when consuming energy, also the fluctuation of green energy production and energy transportation are important factors.

As elastic operations for VM migration require huge bitrate to be available among datacenters for some time periods, the inter-datacenter network can be based on the optical technology and must provide automated interfaces to set-up and tear down optical connections with the required bitrate. Some works consider optical networks to interconnect datacenters. For instance, the authors in [19] present routing algorithms considering both routing and scheduling and compare energy savings with respect to a scenario where routing and scheduling problems are solved separately.

However, to the best of our knowledge, no work compares the way to compute scheduling considering both energy and communications costs in a single framework. In addition, we focus on solar energy, which is more predictable, and take more advantage of our network capabilities to migrate workload. Regarding our power model, we rely on using the Power Usage Effectiveness (PUE) ratio [20], where the consumed power becomes the computational power plus all the extra IT power directly derived from the first one. All the above is considered in the Elastic Operations in Federated Datacenter for Performance and Cost Optimization (ELFADO) problem. Solving ELFADO we reach energy consumption reductions up to 52%, outperforming previous works.

The rest of this article is organized as follows. Section 2 describes a power model for the fat-tree intra-datacenter architecture and presents the motivation of this work: to tackle the ELFADO problem. Two approaches for solving the ELFADO problem are presented: distributed and centralized. In Section 3, the ELFADO problem is formally stated and mathematical models and heuristics algorithms to solve it for both, the distributed and the centralized approach are presented. Illustrative results are provided in Section 4 for a realistic scenario with five datacenters strategically placed around the globe. Finally, Section 5 concludes the article.

2. Orchestrating federated datacenters

In this section we first present the considered power model to evaluate the energy consumption of each individual datacenter when they are based on the fat-tree architecture. Next, we present the main objective of elastic operations, i.e. minimizing operational costs by taking advantage from available green energy and cheap brown energy.

2.1. Datacenter power model

Two main contributions to the power consumption of a datacenter can be distinguished: (i) the power consumed by IT devices, P_{IT}, which comprises both the servers located in the datacenter as well as the switches employed to interconnect those servers; (ii) the power consumption of the non-IT equipment, $P_{non\text{-}IT}$, such as cooling, power supplies and power distribution systems. Thus, total power consumption of a datacenter can be computed as $P_{DC} = P_{IT} + P_{non\text{-}IT}$. P_{IT} can be easily estimated by counting the number of servers and switches of a datacenter. However, it is difficult to evaluate the power consumption of non-IT devices since it depend on several details and factors which cannot be easily estimated. For instance, the power consumption of the cooling system strongly depends on the geographical location of the datacenter and on the building hosting that datacenter.

An indirect way to estimate a numerical value for $P_{non\text{-}IT}$ is to consider the PUE metric [20]. PUE can be used as a measure of the energy efficiency of a datacenter and quantifies the amount of power consumed by non-IT equipment in that datacenter: $PUE = P_{DC}/P_{IT}$. Therefore, if P_{IT} and PUE can be estimated for a given datacenter, the total power consumed in a datacenter can be computed as $P_{DC} = PUE^{*}P_{IT}$.

Regarding P_{IT}, we can distinguish between the power consumed by the servers and by network equipment. The power consumed by a server, $P_{server}(k)$, depends mainly on the CPU load (k) utilization, expressed as the ratio between the current load and the maximum capacity of the server. According to [21], the power consumption of a server can be estimated as $P_{server}(k) = P_{server\text{-}idle} + (P_{server\text{-}max} - P_{server\text{-}idle})^{*}k$, where $P_{server\text{-}idle}$ and $P_{server\text{-}max}$ represent the power consumed by the server when it is idle and when it operates at its maximum capacity, respectively. The power consumed by network equipment depends on the specific architecture of the datacenter. In this work, we assume the *fat-tree* architecture (Fig. 1), which consists of three switching layers; from top to bottom: *Core*, *Aggregation* and *Edge*.

The lower layers – aggregation and edge – together with the servers are organized in a number of clusters M. In each of these clusters, switches have M interfaces operating at the same bitrate. Each cluster has $M/2$ edge switches and $M/2$ aggregation switches, all with M ports; it constitutes a *bipartite* graph by connecting each edge to every aggregation switch. In each edge switch, $M/2$ ports are connected directly to servers and the other $M/2$ ports are connected to $M/2$ ports of the aggregation switches. Thus, each cluster has $M^2/4$ servers and there are $M^3/4$ servers in total in the datacenter. There are $(M/2)^2$ M-port core switches, each having one port connected to each cluster, whilst each cluster is connected to every core switch.

We consider that clusters are active when one or more servers are loaded; otherwise the complete cluster is turned-off. Then, the power consumption of cluster i, $P^i_{cluster}$, can be estimated as,

$$P^i_{cluster} = a^i \cdot \left(\frac{M}{2} \cdot (P_{agg} + P_{edge}) + \sum_{s=1}^{M^2/4} P_{server}(k^i_s) \right), \qquad (1)$$

where a^i indicates whether the cluster is active and P_{agg} and P_{edge} denote the power consumption of aggregation and edge switches. The power consumption of the IT devices in the datacenter can eventually be computed as follows, where P_{core} denote the power consumption of core switches.

$$P_{IT} = \frac{M^2}{4} \cdot P_{core} + \sum_{i=1}^{M} P^i_{cluster}. \qquad (2)$$

2.2. Minimizing energy expenditures

A first optimization to reduce energy expenditures is to perform consolidation, placing VMs so as to load servers as much as possible and switching off those servers that become unused. To further reduce energy consumption, consolidation can be performed by taking into account clusters structure, and switching on/off clusters as single units. Those servers in switched on clusters without assigned load remain active and ready to accommodate spikes in demand.

In addition, as stated in the introduction, datacenter federations can perform elastic operations, migrating VMs among datacenters aiming at minimizing operational costs by taking advantage from available green energy in some datacenters and off-peak cheap brown energy in others datacenters while ensuring the desired QoE level. We use latency experienced by the users of a service as a measure of QoE level.

We face then, the ELFADO problem, which orchestrates federated datacenters providing optimal VMs placement so as to minimize operational costs. We assume that operational costs are dominated by energy and communications costs, so we focus on specifically minimizing those costs.

Two approaches can be devised to orchestrate federated datacenters (Fig. 2): (i) *distributed* (D in Fig. 2), where scheduling algorithms running inside datacenter resource managers compute periodically the optimal placement for the VMs currently placed in the local datacenter; (ii) *centralized* (C in Fig. 2), where a *federation orchestrator* computes periodically the global optimal placement for all the VMs in the federated datacenters and communicates that computation to each datacenter resource manager. In both approaches, local resource managers interface the rest of datacenters to coordinate VM migration and the SDN controlling the interconnection network to request optical datacenter-to-datacenter connections' set-up and teardown.

To solve the ELFADO problem some data must be available, such an estimation of QoE perceived by the users, the amount of green energy available in each datacenter, the cost of brown energy, among others. QoE can be estimated by a specialized module inside each resource manager [22]. The cost of brown energy comes from the contract each datacenter has with the local power supply company, which varies with the time of day. Finally, the amount of green energy that will be likely available in the next period can be predicted using historical data and weather forecast [23]. Each local resource manager can flood all that data to the rest of resource managers in remote datacenters.

For illustrative purposes, Fig. 3 plots unit brown energy cost, c_d, and normalized availability of green energy, δ_d, for datacenter d as a function of the time of day. Brown energy cost varies with the time showing on-peak and off-peak periods, where energy during on-peak is approximately 40% more expensive than during off-peak periods. Regarding green energy availability, large variations during the day can be expected. In the view of Fig. 3, it is clear

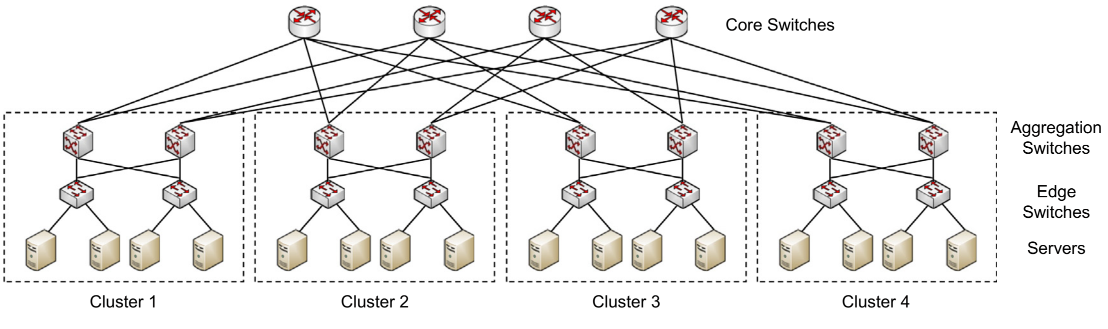

Fig. 1. Example of fat-tree datacenter architecture ($M = 4$).

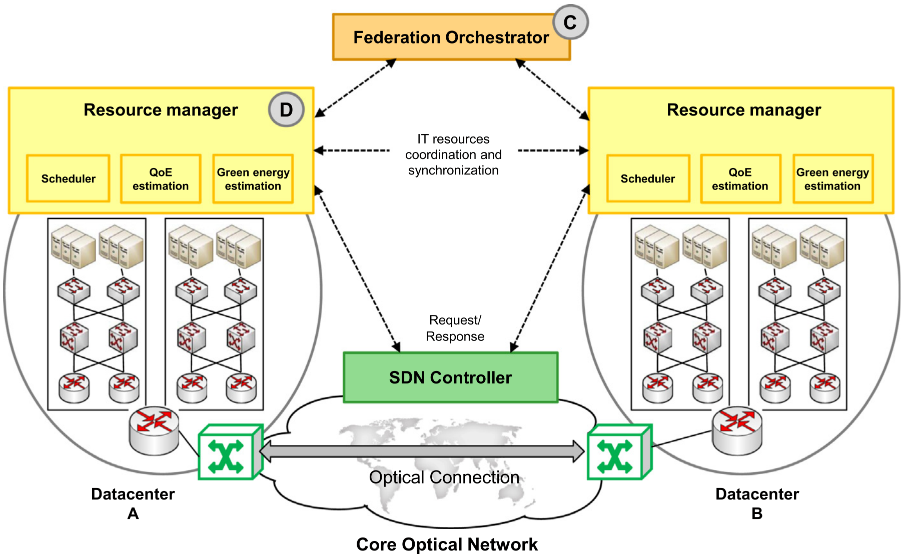

Fig. 2. Distributed and centralized federated datacenters orchestration.

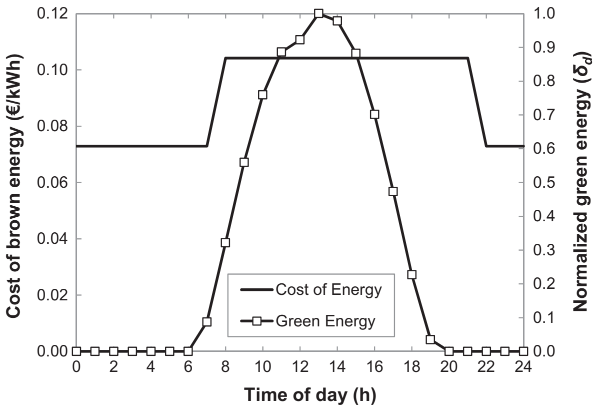

Fig. 3. Unit cost of brown energy and normalized availability of green energy against the time of day.

that some advantage can be taken from orchestrating the federated datacenters, moving VMs to place them in the most advantageous datacenter.

Let us assume that datacenters are dimensioned to cover some proportion β_d of the total energy consumption for the maximum dimensioning. Then, green coverage in datacenter d, α_d, can be estimated as, $\alpha_d(t) = \beta_d * \delta_d(t)$, and the amount of green energy available can be estimated as $g_d(t) = \alpha_d(t) * Energy\_MaxDimensioning$, where *Energy_MaxDimensioning* represents the amount of energy consumed for the maximum dimensioning.

In the distributed approach, local datacenters do not know the amount of VMs that will be placed in each datacenter in the next period, since that decision is to be taken by each datacenter resource manager in the current period. Therefore, the amount of VMs that can take advantage from green energy availability in each datacenter in the next period cannot be computed. To overcome that problem, estimation on the unitary energy cost in each datacenter should be made. We use Eq. (3),

$$\hat{c}_d = (1 - \alpha_d) \cdot c_d, \qquad (3)$$

i.e. the cost of the energy in each datacenter is estimated by decrementing the cost of brown energy with the expected green coverage value. As an example, the estimated cost of the energy is 0.0729 €/kWh at 2 am and 0 €/kWh at 1 pm (assuming $\beta_d = 1$).

In general, however, green energy covers only partially, even in the generation peak, total energy consumption, thus $\beta_d < 1$. Therefore, if several datacenters take the decision of migrating local VMs to one remote datacenter in the hope of reducing costs, it may happen that some brown energy need to be drawn from the grid if not enough green energy is available, which may result in higher energy cost in addition to some communication cost.

In contrast, the amount of VMs to be placed in each datacenter in the next period is known in the centralized approach since the placing decision is taken in the centralized federation orchestrator. Therefore, one can expect that better VM placements can be done in the centralized approach, which might result into further cost savings.

Next section formally state the ELFADO problem and present ILP models and heuristic algorithms for solving efficiently both, the distributed and the centralized approaches.

3. The elfado problem

3.1. Problem statement

The ELFADO problem can be formally stated as follows:
Given:

- a set of federated datacenters D.
- the set of optical connections E that can be established between two datacenters,
- a set of VMs $V(d)$ in each datacenter d,
- a set of client locations L, where n_l is the number of users in location l to be served in the next period,
- PUE_d, brown energy cost c_d, and green coverage level α_d in datacenter d for the next period,
- the data volume k_v and the number of cores $cores_v$ of each VM v,
- energy consumption of each server as a function of the load k, $w_{server}(k) = P_{server}(k) * 1$ h,

- the performance p_{ld} perceived in location l when served from a virtual machine placed in datacenter d,
- a threshold th_v for the performance required at any time for accessing the service in virtual machine v.

Output: the datacenter where each VM will be placed the next time period;

Objective: Minimize energy and communications cost for the next time period ensuring the performance objective for each service.

As previously stated, the ELFADO problem can be solved assuming either a distributed or centralized approach. In the subsequent subsections we present mathematical programming formulations for each of the approaches. In addition, in view of the stringent times required, their exact solving becomes impractical and, as a result, heuristic algorithms are needed so as to provide good near optimal solutions in the time periods required for on-line DC operation.

3.2. Mathematical formulations

The following sets and parameters have been defined:

Sets:

D	set of federated datacenters, index d
E	set of optical connections that can be established, index e
$E(d_1)$	set of optical connections between d_1 and any other datacenter
V	set of virtual machines, index v
$V(d_1)$	set of virtual machines in datacenter d_1
L	set of client locations, index l

Users and performance:

p_{ld}	performance perceived in location l when accessing datacenter d
n_l	number of users in location l
th_v	the threshold performance to be guaranteed for v

Datacenter architecture and VMs:

M	maximum number of clusters per datacenter
n_{server}	number of cores per server
k_v	size in bytes of VM v
n_v	number of cores needed by VM v

Energy:

α_d	green energy cover in datacenter d
g_d	amount of green energy available in datacenter d
PUE_d	PUE for datacenter d
c_d	brown energy cost per kWh in datacenter d
w_v	energy consumption of VM v. It can be computed assuming that the server where it is placed is fully loaded, so $w_v = w_{server\_max}/n_v$

Connections:

k_e	maximum amount of bytes to transfer without exceeding the maximum capacity assigned in connection e. k_e includes the needed overhead from TCP/IP downwards to the optical domain
c_e	cost per Gb transmitted through connection e

Additionally, the decision variables are:

x_{vd}	binary, 1 if virtual machine v is placed in datacenter d, 0 otherwise
y_d	real positive, energy consumption in datacenter d
z_e	integer positive, bytes to transfer through optical connection e

The ILP formulation for the ELFADO problem assuming the distributed approach is as follows. It is worth highlighting that this problem is solved by each of the datacenters separately; in the model, d_1 identifies the local datacenter.

$$(\textit{Distributed ELFADO}) \quad \text{minimize} \quad \sum_{d\in D}(1-\alpha_d)\cdot c_d\cdot y_d + \sum_{e\in E(d_1)} 8\cdot c_e\cdot z_e \tag{4}$$

subject to:

$$\frac{1}{\sum_{l\in L} n_l}\cdot\sum_{l\in L}\sum_{d\in D} n_l\cdot p_{ld}\cdot x_{vd} \leqslant th_v \quad \forall v\in V(d_1) \tag{5}$$

$$\sum_{d\in D} x_{vd} = 1 \quad \forall v\in V(d_1) \tag{6}$$

$$y_d = PUE_d\cdot\sum_{v\in V(d_1)} w_v\cdot x_{vd} \qquad \forall d\in D \tag{7}$$

$$z_{e=(d_1,d_2)} = \sum_{v\in V(d_1)} k_v\cdot x_{vd_2} \quad \forall d_2\in D\setminus\{d_1\} \tag{8}$$

$$z_e \leqslant k_e \quad \forall e\in E(d_1) \tag{9}$$

The objective function in Eq. (4) minimizes the total cost for the VMs in a given datacenter d_1, which consists on the estimated energy costs plus the communication costs for the VMs that are moved to remote datacenters.

Constraint (5) guarantees that each VM is assigned to a datacenter if the on-average performance perceived by the users is above the given threshold. Constraint (6) ensures that each VM is assigned to one datacenter. Constraint (7) computes the energy consumption in each datacenter as a result of moving VM from the local datacenter. Constraint (8) computes the amount of data to be transfer from the local to each remote datacenter. Finally, constraint (9) assures that the capacity of each optical connection from the local datacenter is not exceeded.

The ILP formulation for the centralized one is presented next. Although the model is similar to the distributed approach, this problem computes a global solution for all the datacenters and as a result, of the total amount of VMs that will be placed in the next period in each datacenter can be computed. Therefore, the centralized ELFADO computes the cost of the energy in each datacenter given the amount of green energy available.

Two additional decision variables are defined:

γ_d	positive integer with the number of servers operating with some load in datacenter d
ρ_d	positive integer with the number of clusters switched on in datacenter d

$$(\textit{Centralized ELFADO}) \quad \text{minimize} \quad \sum_{d\in D} c_d\cdot y_d + \sum_{e\in E} 8\cdot c_e\cdot z_e \tag{10}$$

subject to:

$$\frac{1}{\sum_{l\in L} n_l} \cdot \sum_{l\in L}\sum_{d\in D} n_l \cdot p_{ld} \cdot x_{vd} \leqslant th_v \quad \forall v \in V \tag{11}$$

$$\sum_{d\in D} x_{vd} = 1 \quad \forall v \in V \tag{12}$$

$$\gamma_d \geqslant \frac{1}{n_{server}} \cdot \sum_{v\in V} n_v \cdot x_{vd} \quad \forall d \in D \tag{13}$$

$$\rho_d \geqslant \frac{4}{M^2} \cdot \gamma_d \quad \forall d \in D \tag{14}$$

$$\begin{aligned} y_d \geqslant\ & PUE_d \\ & \cdot \Big(\tfrac{M^2}{4} \cdot w_{core} + \tfrac{M}{2} \cdot (w_{agg} + w_{edge}) \cdot \rho_d + w_{server-\max} \cdot \gamma_d + w_{server-idle} \\ & \cdot \Big(\tfrac{M^2}{4} \cdot \rho_d - \gamma_d \Big) \Big) - g_d \quad \forall d \in D \end{aligned} \tag{15}$$

$$z_{e=(d_1,d_2)} = \sum_{v\in V(d_1)} k_v \cdot x_{vd2} \quad \forall d_1, d_2 \in D,_{d_1 \neq d_2} \tag{16}$$

$$z_e \leqslant k_e \quad \forall e \in E \tag{17}$$

The objective function (10) minimizes the total cost for all datacenters in the federation, which consists on the energy costs plus the communication costs for the VMs that are moved between datacenters.

Constraint (11) guarantees that each VM is assigned to a datacenter if the on-average performance perceived by the users is above the given threshold. Constraint (12) ensures that each VM is assigned to one datacenter. Constraint (13) computes, for each datacenter, the amount of servers where some VM is to be placed, whereas constraint (14) computes the number of clusters that will be switched on. Constraint (15) computes the brown energy consumption in each datacenter as the difference between the effective energy consumption, computed as Eqs. (1), (2), and the amount of green energy available in the next period in each datacenter. Note that $w_{(\cdot)}$=$P_{(\cdot)}$ * 1 h. Constraint (16) computes the amount of data to be transfer from each datacenter to some other remote datacenter. Finally, constraint (17) assures that the capacity of each optical connection is not exceeded.

The ELFADO problem is *NP-hard* since it is based upon the on the well-known capacitated plant location problem which has been proved to be *NP-hard* [24]. Regarding problem sizes, the number of variables and constraints for each approach are detailed in Table 1. Additionally, an estimation of problems' size is calculated for the scenario presented in Section 4.

Although the size of the ILP models is limited, they must be solved in real time (in the order of few seconds). In our experiments described in Section 5, we used commercial solvers such as CPLEX [25] to solve each approach. The distributed approach took tens of minutes on average to be solved; more than 1 h in the worst case, whereas the centralized approach took more than one hour on average. As a consequence, in the next section we propose heuristic algorithms that provide much better trade-off between optimality and complexity to produce solutions in practical computations times, short enough to be used for schedule real federated datacenters.

Table 1
Size of the ELFADO problem.

	Constraints	Variables
Distributed	O(|V| + |D|) (10^4)	O(|V| · |D|) (10^5)
Centralized	O(|V| + |D|^2) (10^5)	O(|V| · |D|) (10^5)

3.3. Heuristic algorithms

The heuristic algorithm for the distributed approach (Algorithm 1) schedules the set of VMs in the local datacenter. For each VM, all feasible, in terms of performance (p_{vd}), placements are found and the cost for that placement is computed (lines 2–9). If the placement is in the local datacenter, only energy costs are considered, whereas if it is in a remote datacenter communication costs are also included. Note that energy costs are estimated using the green energy cover to decrement the cost of the energy in the considered datacenter. The list of feasible placements is ordered as a function of the cost (line 10). Each VM is placed afterwards in the cheapest datacenter provided that the amount of data to be transferred through the optical connection does not exceed the maximum available, in case of a remote placement (lines 11–17). The final solution is eventually returned (line 18).

Algorithm 1 Heuristic for the distributed ELFADO.

```
INPUT d_1, V(d_1), D
OUTPUT Sol
1:  Sol ← Ø
2:  for each $v \in V(d_1)$ do
3:    for each $d \in D$ do
4:      if $p_{vd} \leqslant th_v$ then
5:        if $d \neq d_1$ then
6:          let $e = (d_1, d)$
7:          $C[v]$.list ← {d, e, $(1 - \alpha_d) * c_d * w_v + c_e * k_v$}
8:        else
9:          $C[v]$.list ← {d, Ø, $(1 - \alpha_d) * c_d * w_v$}
10:     sort ($C[v]$.list, Ascending)
11:   for each $v \in V(d_1)$ do
12:     for $i$ = 1..$C[v]$.list.length do
13:       $\{d,e\}$ ← $C[v]$.list($i$)
14:       if $e \neq$ Ø && $z_e + k_v > k_e$ then continue
15:       if $e \neq$ Ø then $z_e$ ← $z_e + k_v$
16:       Sol ← Sol ∪ {$(v,d)$}
17:       break
18:   return Sol
```

The heuristic algorithm for the centralized approach (Algorithm 2) schedules the set of VMs in all the federated datacenters. The proposed heuristic focuses on taking advantage from all the available green energy, only considering the cost of brown energy and communications when no more green energy is available. The perceived performance of each VM in its current placement is computed; those infeasible placements (the perceived performance is under the threshold) are added to set *U* whereas those which are feasible to the set *F* (lines 2–7). Next, the remaining green energy in each datacenter is computed, considering the available green energy and the energy consumption of those feasible placements (line 8). The set *R* stores those datacenters with remaining green energy available.

The remaining green energy in the datacenters (if any) is used to place infeasible placements in set *U*; the cheapest feasible placement if found for each VM in *U* provided that the energy consumption of that VM can take advantage from remaining green energy (lines 12–15). If a feasible placement is finally found, the remaining green energy for the selected datacenter is updated (line 16) and if no green energy remains available, that datacenter is eventually removed from set *R*. The same process of maximizing available green energy is performed for the feasible placements in set F (lines 19–25).

Every remaining not yet considered, feasible or unfeasible, placement is stored in the set F to be jointly considered (line 26) and an algorithm similar to the one for the distributed approach is then followed (lines 27–42). The only difference is that the cost of new placements is computed considering that all the energy will come from brown sources (lines 32 and 34). Finally, the solution for all the datacenters is returned.

Algorithm 2 Heuristic for the centralized ELFADO.

```
INPUT $V$, $D$
OUTPUT $Sol$
1:  Initialize $Sol \leftarrow Ø$; $U \leftarrow Ø$; $F \leftarrow Ø$; $R \leftarrow Ø$
2:  for each $d \in D$ do
3:    $U_d \leftarrow Ø$; $F_d \leftarrow Ø$
4:    for each $v \in V(d)$ do
5:      if $p_{vd} > th_v$ then
6:      $U_d \leftarrow U_d \cup \{(v,d)\}$
7:      else $F_d \leftarrow F_d \cup \{(v,d)\}$
8:    $r_d \leftarrow g_d -$ computeEnergy $(F_d)$
9:    $U \leftarrow U \cup U_d$; $F \leftarrow F \cup F_d$
10:  if $r_d < 0$ then
11:    $R \leftarrow \{(d,r_d)\}$
12:  if $R \neq Ø$ then
13:    for each $(v,d_1) \in U$ do
14:      find $(d_2,r_{d2}) \in R$ feasible for $v$ such that $r_d > w_v$ with
         min comm cost
15:      $Sol \leftarrow Sol \cup \{(v,d_2)\}$
16:      $r_{d2} \leftarrow r_{d2} - \mathrm{PUE}_{\mathrm{d2}} * w_v$
17:      if $r_{d2} < =0$ then
18:        $R \leftarrow R \setminus \{(d_2,r_{d2})\}$
19:  if $R \neq Ø$ then
20:    for each $(v,d_1) \in F$ do
21:      find $(d_2, r_{d2}) \in R$ feasible for $v$ such that $r_d > w_v$ with
         min comm cost
22:      $Sol \leftarrow Sol \cup \{(v,d_2)\}$
23:      $r_{d2} \leftarrow r_{d2} - \mathrm{PUE}_{d2}^{*} w_v$
24:      if $r_{d2} < =0$ then
25:        $R \leftarrow R \setminus \{(d_2, r_{d2})\}$
26:  $F \leftarrow F \cup U$
27:  for each $\{v,d_1\} \in F$ do
28:    for each $d_2 \in D$ do
29:      if $p_{vd2} \leqslant th_v$ then
30:        if $d_2 \neq d_1$ then
31:          let $e = (d_1,d_2)$
32:          $C[v]$.list $\leftarrow (d_2, e, c_{d2} * w_v + c_e * k_v)$
33:      else
34:          $C[v]$.list $\leftarrow (d_2, e, c_{d2} * w_v)$
35:    sort $(C[v]$.list, Ascending)
36:  for each $(v,d_1) \in F$ do
37:    for $i = 1..C[v]$.list.length do
38:      $(d_2,e) \leftarrow C[\mathrm{v}]$.list$(i)$
39:      if $e \neq Ø$ && $z_e + k_v > k_e$ then continue
40:      if $e \neq Ø$ then $z_e \leftarrow z_e + k_v$
41:      $Sol \leftarrow Sol \cup \{(v, d_2)\}$
42:      break
43:    return $Sol$
```

The performance of each of the proposed heuristic algorithms was compared against the corresponding ILP model. In all the experiments performed, the heuristics were able to provide a much better trade-off between optimality and computation time; in all the tests the optimal solution was found within running times of hundreds of milliseconds, in contrast to tens of minutes (for the distributed) and even hours (for the centralized) needed to find the optimal solution with the ILP models. Thus, we use the heuristics to solve the instances in the scenario presented in the next section.

4. Performance evaluation

In this section, we present the scenario considered in our experiments and we show the results from solving the ELFADO problem considering a realistic instance; we evaluate the impact in the cost when distributed and centralized approaches are used for scheduling VM placement compared to a fixed placement, where no scheduling is done.

4.1. Scenario

We implemented the proposed heuristic algorithms for the distributed and centralized ELFADO approaches on a scheduler in the OpenNebula cloud management middleware [26]. For comparison, a *fixed* approach, where the total workload is evenly distributed among the federated datacenters, was also implemented.

We consider the global 11-location topology depicted in Fig. 4. Each location collects user traffic towards the set of federated datacenters, which consists of five datacenters strategically located in Taiwan, India, Spain, and Illinois and California in the USA. A global telecom operator provides optical connectivity among datacenters, which is based upon the flexgrid technology. The number of users in each location was computed considering Wikipedia's audience by regions [27] that was scaled and distributed among the different locations in each region. Latency between location pairs was computed according to [22].

Table 2 briefly presents the value considered for some representative energy parameters. Daily PUE values were computed according to [15] using data obtained from [28]. Green energy coverage was obtained from [28–30] and brown energy cost for each datacenter was estimated from their respective local electric company rates (e.g. [31], [32]). Servers in datacenters are assumed to be HP ProLiant DL580 G3,[1] equipped with four processors, 2 cores per processor, with $P_{server\text{-}idle} = 520$ W and $P_{server\text{-}max} = 833$ W.

In line with [14], datacenters are dimensioned assuming a fat-tree topology with a maximum of $M = 48$ clusters with two levels of switches and $M^2/4 = 576$ servers each. The number of VMs was set to 35,000, with individual image size of 5 GB; we assume that each VM runs in one single core. An integer number of clusters is always switched on, so as to support the load assigned to the datacenter; those servers without assigned load remain active and ready to accommodate spikes in demand. Green cover was set to ensure, at the highest green energy generation time, a proportion of energy β_d when all VMs run in datacenter d.

We consider a different type of switch, and thus a different power consumption value, for each layer of the intra-datacenter architecture. We selected the Huawei[2] CloudEngine switches series; Table 3 specifies model, switching capacity and power consumption for each considered model.

Finally, we consider that each datacenter is connected to the flexgrid inter-datacenter network through a router equipped with 100 Gb/s bandwidth variable transponders. Therefore, the actual capacity of optical connections is limited to that value. To compute the real throughput, we consider headers for the different protocols, i.e. TCP, IP, and GbE. The maximum amount of bytes to

[1] Hewlett-Packard, http://www.hp.com/.

[2] Huawei, http://www.huawei.com.

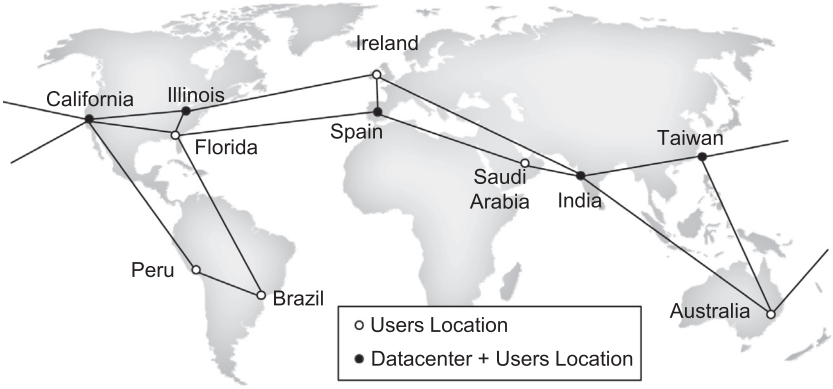

Fig. 4. Considered scenario: federated datacenters, locations and inter-datacenter network.

Table 2
Value of energy parameters.

Datacenter	c_d (on/off peak) (€/kWh)	β_d	PUE (max/avg)
Taiwan	0.0700/0.0490	0.5	1.671/1.632
India	0.0774/0.0542	0.9	1.694/1.694
Spain	0.1042/0.0729	0.9	1.670/1.457
Illinois	0.0735/0.0515	0.2	1.512/1.368
California	0.0988/0.0692	0.5	1.385/1.303

Table 3
Characteristics of Huawei CloudEngine switches.

Layer	Model	Sw. capacity	Power consumption
Core	12,812	48 Tb/s	P_{core} = 16,200 W
Aggregation	6800	1.28 Tb/s	P_{agg} = 270 W
Edge	5800	336 Gb/s	P_{edge} = 150 W

transfer, k_e, was computed to guarantee that VM migration is performed in less than 40 min.

4.2. Performance evaluation

Fig. 5 (left) shows the availability of green energy as a function of the time (GMT) at each datacenter, $\alpha_d(t)$, for a typical spring day, whereas the two rightmost graphs in Fig. 5 illustrate the behavior of the distributed (center) and centralized (right) ELFADO approaches.

The distributed approach places VMs in datacenters where the cost of energy (plus communications) is expected to be minimum in the next period; Eq. (3) is used for that energy cost estimation. However, in view of Fig. 5 (center), it is clear that Eq. (3) does not provide a clear picture, since all VMs are placed in India and Spain during the day periods where more green energy is available in those locations, thus exceeding green energy availability and paying a higher cost. In contrast, datacenter in Illinois seems to be very little utilized.

Interestingly, the centralized approach reduces the percentage of VMs in those datacenters with higher green coverage, to place only the amount of VMs (translated into powered-on clusters and servers) that the available green energy can support and placing the rest considering brown energy (and communications) costs. In fact, the datacenter in Illinois is more used in the centralized approach as a consequence of its cheaper brown energy cost compared to that of California.

Fig. 6 presents costs and performance as a function of the time for all three approaches; cost per transmitted bit was set to 1e − 9 €/Gb*km. Energy costs per hour plots in Fig. 6 (left) show a remarkable reduction in energy costs when some ELFADO approach is implemented, with respect to the fixed approach. Daily comparison presented in Table 4 show savings of 11% for the distributed and over 52% for the centralized approach. Hourly plot for the distributed approach clearly highlight how by placing VMs in datacenters where the estimated energy is cheaper results in a high amount brown energy being drawn from the grid at a more expensive price. In contrast, the centralized approach leverages green energy arriving to virtually zero energy cost in some periods.

Regarding communications (Fig. 6 (center)), the distributed approach shows a more intensive use, presenting three peaks, exactly when the datacenter in Illinois is used to compensate energy costs between green energy availability peaks in the rest of datacenters. However, although the centralized approach is less communications intensive, the total daily communications costs are only under 6% cheaper compared to the distributed approach, as shown in Table 4.

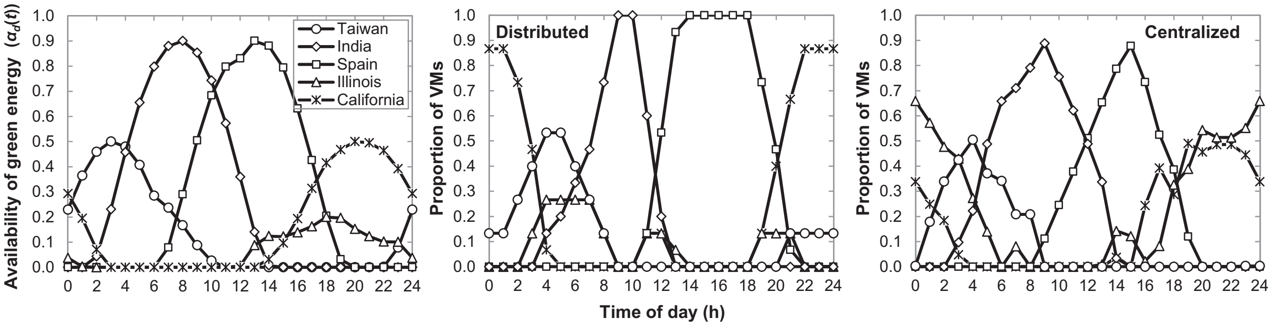

Fig. 5. Availability of green energy vs. time in all datacenters (left). Percentage of VMs in each datacenter when the distributed (center) and the centralized (right) approaches are applied.

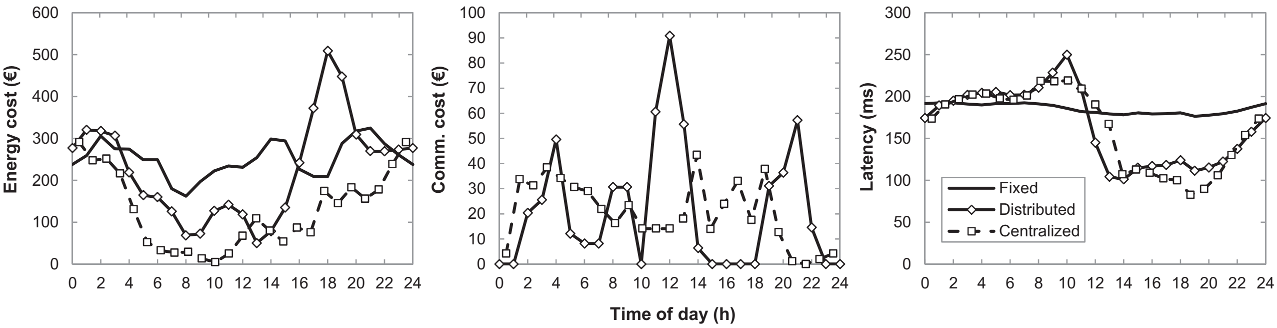

Fig. 6. Energy (left) and communication (center) cost per hour against time. Latency vs. time (right).

Table 4
Comparison of daily costs and performance.

Approach	Energy cost	Comm. cost	Total cost	Average latency
Fixed	6048 €	–	6048 €	185.2 ms
Distributed	5374 € (11.1%)	537 €	5912 € (2.3%)	164.2 ms (11.3%)
Centralized	2867 € (52.6%)	508 € (5.8%)	3376 € (44.2%)	161.5 ms (12.8%)

Aggregated daily costs are detailed in Table 4 for all three approaches. As shown, the distributed approach saves only 2% of total cost when compared to the fixed approach. Although, that percentage represents more than 100€ per day, it is just a small portion of the savings obtained by the centralized approach, which are as high as just over 44% (more than 2.6 k€ per day).

Regarding performance (latency), both the distributed and the centralized approach provide figures more than 10% lower than that of the fixed approach as shown in Table 4. Hourly plots presented in Fig. 6 (right) show that latency is slightly higher during some morning periods under both, the distributed and the centralized ELFADO, with respect to that of the fixed; after noon, however, both approaches reduce latency extraordinary since VMs are placed closer to users.

The results presented in Fig. 6 were obtained by fixing the value of th_v to 1.3 * average(*latency_fixed*) (specified in Table 4), so as to allow obtaining worse hourly performance values in the hope of obtaining better daily ones. Fig. 7 gives insight of the sensitivity of costs to the value of that threshold. Fixed costs are also plotted as a reference. Costs in the centralized approach show that even for very restrictive thresholds, noticeable cost savings can be obtained. In addition, when the threshold is set to the average latency in the fixed approach or above, obtained costs are almost constant. In contrast, the distributed approach proves to be more sensible to that threshold, reaching a minimum in terms of costs when the threshold value is 30% over the average latency in the fixed approach.

Finally, Fig. 8 illustrates the influence of the cost per bit to transfer VMs from one datacenter to another. As before, fixed costs are plotted for reference. Energy costs in the distributed approach increase sharply when the cost per bit doubles, almost preventing

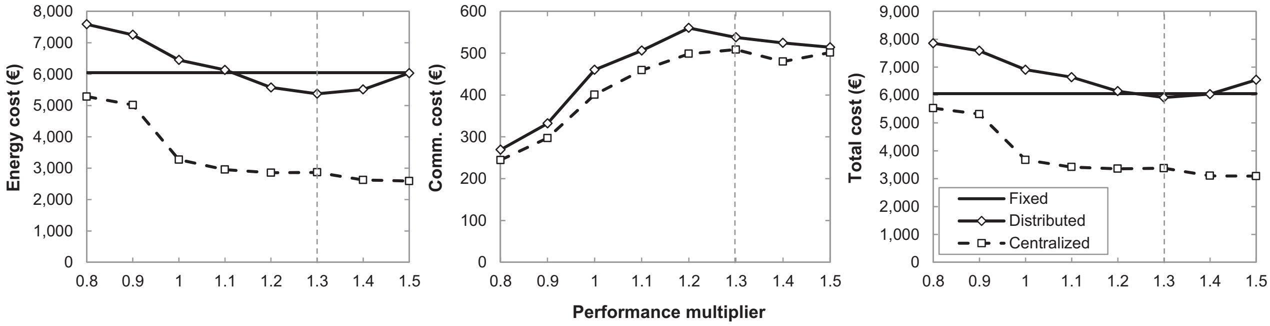

Fig. 7. Cost per day vs. performance threshold.

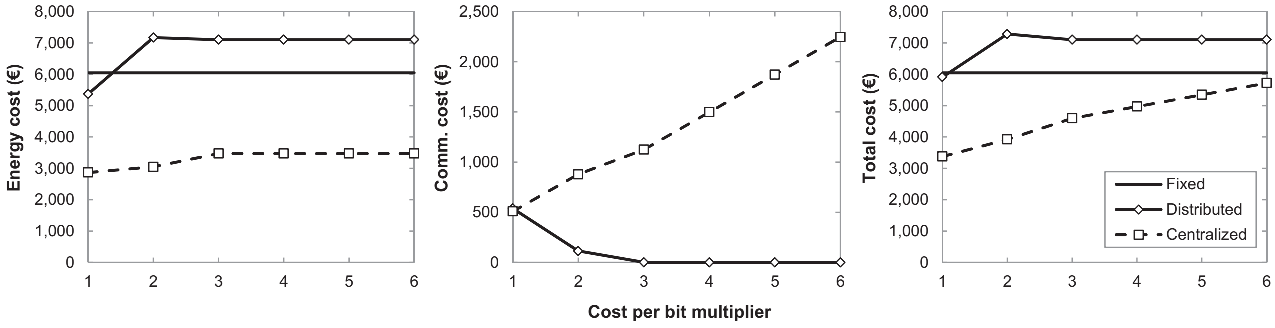

Fig. 8. Cost per day against cost per bit.

from moving VMs, as clearly shown Fig. 8 (center). Nonetheless, energy costs are almost stable in the centralized approach. Recall that the proposed heuristic focuses on green energy availability as the first indicator for placing VMs. In fact, communication cost increase linearly with the increment in the cost per bit. However, it is not until the cost per bit increases more than 6 times when the centralized approach cost equals that of the fixed approach.

5. Concluding remarks

The enormous energy consumption of datacenters translates into high operational expenditures for datacenter operators. Although the use of green energy allows reducing energy bill, its availability is reduced depending on the hour of the day, weather and season, etc. Federating datacenters can be a way for independent datacenter operators to, not only increase their revenue, but also reduce operational expenditures. Aiming at optimizing costs whilst ensuring the desired QoE for users, this work described and formally stated the ELFADO problem to orchestrate federated datacenters, placing workloads in the most convenient datacenter.

Two approaches to solve the ELFADO problem were compared, distributed and centralized, where mathematical formulations as well as heuristic algorithms for scheduling VM placement were proposed. The distributed approach is based on running scheduling algorithms inside datacenter resource managers to compute periodically the optimal placement for the VMs currently in the local datacenter. VMs are placed in datacenters where the cost (energy and communications) is expected to be minimum for the next period. In this approach, energy costs are estimated since the total amount of VMs to be placed in each datacenter is computed in a distributed manner. Therefore, the available green energy could not be enough to cover the whole energy consumption in each datacenter. In contrast, the centralized approach, proposes a *federation orchestrator* to compute the global optimal placement for all the VMs in the federated datacenters. VMs are placed in datacenters so as to take full advantage from green energy availability. This is possible as a result of computing the placement of all VMs in the proposed federation orchestrator at the same time.

The results showed that both ELFADO approaches improve QoE by reducing average latency more than 10% with respect to a fixed approach where no scheduling is performed. Regarding costs, the distributed approach can save up to 11% of costs with respect the fixed approach. However, when communication costs are considered, total cost savings reduce to only 2%. The centralized approach showed remarkable energy cost savings circa 52%, which result in 44% when communication costs are taking into account. Finally, it was shown that the centralized approach provides costs savings even when the cost per bit increases 6 times.

Acknowledgments

The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement no 317999 IDEALIST project and no 257740 TREND NoE. Moreover, it was supported by the Spanish science ministry through the TEC2011-27310 ELASTIC project.

References

[1] M. Armbrust et al., A view of cloud computing, Commun. ACM 53 (2010) 50–58.
[2] US Environmental Protection Agency, Report to Congress on Server and Data Center, Energy Efficiency, 2007.
[3] M. Mishra, A. Das, P. Kulkarni, A. Sahoo, Dynamic resource management using virtual machine migrations, IEEE Commun. Mag. 50 (2012) 34–40.
[4] I. Goiri, F. Julià, R. Nou, J.L. Berral, J. Guitart, J. Torres, Energy-aware scheduling in virtualized datacenters, in: Proc. IEEE International Conference on Cluster Computing, 2010.
[5] X. Zhao, V. Vusirikala, B. Koley, V. Kamalov, T. Hofmeister, The prospect of inter-data-center optical networks, IEEE Commun. Mag. 51 (2013) 32–38.
[6] I. Goiri, J. Guitart, J. Torres, Characterizing cloud federation for enhancing providers' profit, in: Proc. IEEE International Conference on Cloud Computing, 2010.
[7] T. Kudoh, G. Roberts, I. Monga, Network services interface. An interface for requesting dynamic inter-datacenter networks, in: Proc. OSA OFC, 2013.
[8] M. Jinno, H. Takara, B. Kozicki, Y. Tsukishima, Y. Sone, S. Matsuoka, Spectrum-efficient and scalable elastic optical path network: architecture, benefits, and enabling technologies, IEEE Commun Mag. 47 (2009) 66–73.
[9] ITU-T, Architecture for the automatically switched optical network, Rec. 8080, 2012.
[10] D. King, A. Farrel, A PCE-based architecture for application-based network operations, IETF draft, 2014.
[11] L. Velasco, A. Asensio, J. Ll. Berral, V. López, D. Carrera, A. Castro, J.P. Fernández-Palacios, Cross-stratum orchestration and flexgrid optical networks for datacenter federations, IEEE Network Mag. 27 (2013) 23–30.
[12] L. Velasco, A. Asensio, J.L. Berral, A. Castro, V. López, Towards a carrier SDN: an example for elastic inter-datacenter connectivity, OSA Optics Express 22 (2014) 55–61.
[13] Y. Zhang, N. Ansari, On architecture design, congestion notification, TCP incast and power consumption in data centers, IEEE Commun. Surv. Tutorials 15 (2013) 39–64.
[14] M. Al-Fares, A. Loukissas, A. Vahdat, A scalable, commodity data center network architecture, in: Proc. ACM SIGCOMM, 2008.
[15] I. Goiri, et al., GreenHadoop: leveraging green energy in data-processing frameworks, in: Proc EuroSys, 2012.
[16] L. Liu, H. Wang, X. Liu, X. Jin, W. He, Q. Wang, Y. Chen, GreenCloud: a new architecture for green data center, in: Proc. ICAC-INDST, 2009.
[17] Z. Liu, M. Lin, A. Wierman, S. Low, L. Andrew, Geographical load balancing with renewables, in: Proc. of ACM Greenmetrics, 2011.
[18] J.M. Pierson, Green task allocation: taking into account the ecological impact of task allocation in clusters and clouds, J. Green Eng. 1 (2011) 129–144.
[19] J. Buysse, K. Georgakilas, A. Tzanakaki, M. De Leenheer, B. Dhoedt, C. Develder, Energy-efficient resource-provisioning algorithms for optical clouds, IEEE/OSA J. Opt. Commun. Networking 5 (2013) 226–239.
[20] The Green Grid: <www.thegreengrid.org/>.
[21] X. Fan, W. Weber, L.A. Barroso, Power provisioning for a warehouse-sized computer, in: Proc. of ACM International Symposium on Computer Architecture (ISCA), 2007.
[22] Verizon, <http://www.verizonbusiness.com/about/network/latency/>.
[23] N. Sharma, et al., Cloudy computing: leveraging weather forecasts in energy harvesting sensor systems, in: Proc. SECON, 2010.
[24] J. Díaz, E. Fernández, A branch-and-price algorithm for the single source capacitated plant location problem, J. Oper. Res. Soc. (JORS) 53 (2002) 728–740.
[25] CPLEX, <http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/>.
[26] OpenNebula, <http://www.opennebula.org/>.
[27] Meta-Wiki. <http://meta.wikimedia.org/wiki/User:Stu/comScore_data_on_Wikimedia#Geographic_breakdown>.
[28] US Department of Energy. US Energy Information Administration. 2009. Web site <http://www.eia.doe.gov/>.
[29] US Department of Energy, <http://apps1.eere.energy.gov/buildings/energyplus/weatherdata_about.cfm>.
[30] D. King, W. Boyson, J. Kratochvil, Photovoltaic Array Performance Model, Sandia National Laboratories, Report, SAND2004-3535, 2004.
[31] Europe's Energy Portal: <http://www.energy.eu/>.
[32] U.S. Department of Labor, Average Energy Prices, <http://www.bls.gov/ro5/aepchi.htm>.