Dynamic VM Placement Method for Minimizing Energy and Carbon Cost in Geographically Distributed Cloud Data Centers

Atefeh Khosravi, *Member, IEEE*, Lachlan L. H. Andrew, *Senior Member, IEEE*, and Rajkumar Buyya, *Fellow, IEEE*

Abstract—Cloud data centers consume a large amount of energy that leads to a high carbon footprint. Taking into account a carbon tax imposed on the emitted carbon makes energy and carbon cost play a major role in data centers' operational costs. To address this challenge, we investigate parameters that have the biggest effect on energy and carbon footprint cost to propose more efficient VM placement approaches. We formulate the total energy cost as a function of the energy consumed by servers plus overhead energy, which is computed through power usage effectiveness (PUE) metric as a function of IT load and outside temperature. Furthermore, we consider that data center sites have access to renewable energy sources. This helps to reduce their reliance on "brown" electricity delivered by off-site providers, which is typically drawn from polluting sources. We then propose multiple VM placement approaches to evaluate their performance and identify the parameters with the greatest impact on the total renewable and brown energy consumption, carbon footprint, and cost. The results show that the approach which considers dynamic PUE, renewable energy sources, and changes in the total energy consumption outperforms the others while still meeting cloud users' service level agreements.

Index Terms—Cloud computing, green computing, energy consumption, data centers, VM placement

1 Introduction

Cloud computing is considered a big step towards the long held dream of delivering computing as a utility to users [1]. The cloud enables access to hardware resources, infrastructure, and software anytime and anywhere on a pay-as-you-go model. Services by cloud computing are delivered by data centers that are distributed across the world, which can host small numbers to thousands of servers. A major issue with these data centers is that they consume a large amount of energy. According to a report from NRDC [2], US data centers power consumption estimation alone in 2013 was 91 billion kilowatt-hours of electricity. This is equivalent to two years' power consumption of New York City's households and is estimated to increase to 140 billion kilowatt hours by 2020, which is responsible for emission of nearly 150 million tons of carbon pollution.

The high energy consumption by data centers incurs high costs to cloud providers, since energy related costs are the most significant cost for a data center [3]. Furthermore, to enforce the environmental sustainability, some countries set carbon tax on the emitted CO_2 [4]. Therefore, monitoring the amount of energy consumed by a data center and the source of the energy, which directly affects the carbon footprint and carbon tax, helps cloud providers to reduce the energy and carbon cost as a major sector of their total cost.

In this paper, we investigate parameters that affect the total cost associated with the energy consumption and carbon footprint for a cloud provider. Here, we only consider the cost of these two parameters, unless otherwise mentioned. A cloud provider often maintains geographically distributed data center sites, similar to popular cloud providers (e.g., Amazon, Google, and Microsoft). Having several sites not only increases the availability, it also gives the cloud provider the option of choosing the destination site based on different criteria upon the reception of the user request (virtual machine requests in this paper). There are different challenges a cloud provider faces to make the decision regarding VM placement and scheduling. In this paper, we study the selection process among several data center sites. Each data center can get its electricity from different electricity providers, we refer to this as off-site brown energy sources, or even can draw the required electricity from on-site renewable ("green") energy sources, such as solar and wind. Having data center sites that can get their power from renewable sources partially or completely helps the provider decrease its dependency on the electricity drawn from off-site grids, which is costly and less clean. Second, off-site brown energy at different locations have different carbon intensities and carbon taxes. Therefore, by

- *A. Khosravi and R. Buyya are with the Cloud Computing and Distributed Systems (CLOUDS) Laboratory, School of Computing and Information Systems, The University of Melbourne, Parkville VIC 3010, Australia. E-mail: atefehk@student.unimelb.edu.au, rbuyya@unimelb.edu.au.*
- *L.L.H. Andrew is with the Clayton School of Information Technology, Monash University, Melbourne, Clayton VIC 3800, Australia. E-mail: lachlan.andrew@monash.edu.*

Manuscript received 13 July 2016; revised 7 Apr. 2017; accepted 23 May 2017. Date of publication 1 June 2017; date of current version 13 July 2017. (Corresponding author: Atefeh Khosravi.)
Recommended for acceptance by L. Dobre, G. Mastorakis, C.X. Mavromoustakis, and F. Xhafa.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TSUSC.2017.2709980

the change of the availability of renewable energy during the day and in the case that they are not available, cloud provider can select the cleanest source of electricity with less carbon tax. The third advantage of having different energy sources at different locations is changes in electricity price, as we consider variable energy pricing during times of the day, i.e., on-peak and off-peak prices.

The last and one of the most important parameters that affects data centers energy consumption, carbon footprint, and their associated cost is the overhead power, e.g., power supplies, cooling, lightning, and UPS. The metric used to demonstrate the overhead is known as Power Usage Effectiveness (PUE) that is defined by The Green Grid consortium [5]. PUE is equal to the data center's total power consumption, which is the input power that goes to the data center, divided by the IT devices power consumption (PUE=TotalPower/ITDevicesPower). If PUE is equal to 1 it means that the data center is perfectly efficient, which is not practically attainable. An increase in PUE indicates more waste of power to support IT devices in the data center. Although state of the art cloud-scale data centers can achieve a PUE of 1.1 [6] or 1.2 [7], cloud providers often collocate with smaller data centers, which can still have PUEs up to 2 [8]. To increase a data center's efficiency, we should identify variables that have the highest impact on the increase of the system's overhead power. The main variable that affects efficiency and PUE value is IT load. When the IT load is increased, CPUs perform in higher frequencies and servers consume more power. This leads to increase in data center's overall load and inside temperature; accordingly the need arises for more power for the cooling of the infrastructure. The second important variable that affects PUE is the outside temperature, which has a great effect on the cooling system power consumption. As outside temperature increases, the data center needs to use the chillers along with the computer room air handler (CRAH), which leads to a significant increase in the power consumption and PUE value. We exploit a model for PUE as a function of IT load and outside temperature and perform VM placement based on dynamic changes of PUE.

The *key contribution* of this paper is a new method for the initial placement of VMs in geographically distributed cloud data centers that simultaneously considers the cost of 1) overhead energy 2) servers' energy and 3) carbon footprint. Moreover, the proposed VM placement method maximizes renewable energy utilization at each data center to minimize the total cost. Finally, we present efficient two-stage VM placement approaches that respond to dynamic PUEs. We also present variations of our method, which explore the effects of different parameters in minimizing energy and carbon cost for a cloud computing environment. To achieve this, we have carried out the following:

- Developed an analytical model of the total cost incurred by the energy consumption and carbon footprint for the data centers.
- Modeled PUE as a function of IT load and outside temperature to incorporate overhead energy consumption, e.g., power supplies, cooling, lightning, and UPS, along with the energy consumed by the servers.
- Used different carbon intensities and carbon taxes for energy sources at each data center site.
- Analyzed the effect of distributing load among data center sites with access to intermittent renewable energy sources.

The reminder of the paper is organized as follows: In Section 2, the related work is discussed. Section 3 discusses the system model, parameters, objective function and constraints. The proposed VM placement approaches are discussed in Section 4. The experimental environment and the performance analysis of the proposed VM placement approaches are presented in Section 5. Finally, Section 6 concludes the paper and presents future directions.

2 Related Work

Over the last few years, there have been extensive studies on reducing energy consumption of cloud data centers. Recently, there has been much interest in reducing data center carbon footprints and energy consumption due to the environmental concerns (specifically around global warming), social pressure, and the prospect of a carbon tax. Most of the early work focuses on making a single server energy efficient by considering hardware aspects and using techniques such as CPU DVFS (dynamic voltage and frequency scaling) [9], [10]. Moreover, virtualization [11] as the foundation of cloud computing systems, enables consolidation [12] and VM migration [13]. There is ongoing research on the later techniques, but the main issue is that they are reactive and require resume and suspension of VMs which cause overhead to the system [14]. Therefore, these techniques should be applied only when they are cost-effective. Lin et al. [15] and Shen et al. [16] used a pro-active technique, known as dynamic right-sizing, to predict the number of active servers needed to host the incoming workload. Since idle servers consume almost half of the peak power [17], this technique could reduce the energy consumption significantly. Lefevre et al. [18] proposed an advanced resource reservation architecture to have a better prediction of the incoming requests by users. The above-mentioned techniques are in the scope of a single data center and they only consider the aspect of reducing energy consumption; which does not necessarily lead to a reduction in the carbon footprint. Aksanli et al. [19] use predication-based algorithms to maximize the usage of renewable energy sources and in the meantime minimize the number of canceled jobs.

One of the first works to reduce cost and brown energy consumption by load distribution among several data center sites, is that of Le et al. [20]. Their work is based on considering the electricity price and energy source (green or brown) to calculate the number of requests each data center can host within a specific time period and budget. However, they do not differentiate among sites that have brown energy sources with different carbon emission rates. Further, the incoming workload is based on SaaS (Software as a Service) requests for Internet services with short processing times, usually in milliseconds. Liu et al. [21] consider geographical load balancing to minimize brown energy consumption as well. They use an optimal mix of renewable energy (solar and wind) along with energy storage in data centers to eliminate brown energy consumption. Lin et al. [22] extended the previous work to find the best estimate combination for solar and wind energy while having net-zero brown energy usage. The MinBrown workload scheduling algorithm is proposed by Chen et al. [23] to minimize brown energy consumption. This algorithm forwards the incoming request to all data centers, then based on the request deadline and brown energy consumption schedules request for execution. Celesti et al. [24] proposed a federated CLEVER-based cloud environment; which is based on allocation of VM requests to the cloud data center with the highest amount of solar energy and lowest cost. Le et al. [25] proposed an optimization-based framework to minimize brown energy consumption and leverage green energy through distribution of the Internet services to the data centers, considering different electricity prices, data center location with different time zones, and access to green energy sources.

Le et al. [26] apply dynamic load distribution policies and cooling strategies to minimize the overall cloud provider's cost but places no cost on carbon emissions. Their work is based on intelligent placement of VM requests on data centers considering the geographical location, time zone, energy price, peak power charges, and cooling system energy consumption. Ren et al. [27] proposed a provably-efficient on-line algorithm (GreFar) with the objective to minimize energy cost. They use servers' energy efficiency information and locations with low electricity prices to schedule batch jobs and, if necessary, suspend the job and resume later. Work by Goiri et al. [28] aims to find the best place for a data center, based on geographical location and data center characteristics to minimize cost, energy consumption, and carbon footprint. Garg et al. [29] proposed an environment-conscious meta-scheduler to minimize carbon emission and maximize cloud provider's profit. They used near-optimal scheduling policies to send HPC (high performance computing) applications to the data center with the least carbon emission and maximum profit, considering applications deadline. They also address the issue of energy consumption and carbon footprint by proposing a novel green cloud architecture [30]. This architecture uses two directories so the cloud providers can register their offered services. A notable work by Buchbinder et al. [31] has the same objective of reducing the energy cost of a cloud provider with multiple data center sites. They perform online migration of running batch jobs among data center sites, taking advantage of dynamic energy pricing at different locations, while considering the network bandwidth costs among data centers and future changes in electricity price. Similarly, Giacobbe et al. [32] perform VM migration between cloud data centers participating in a federated environment to push down energy costs. They take advantage of dynamic electricity pricing to migrate the VMs to the data center with lowest energy cost and enough resources. Another work by Giacobbe et al. [33] uses the idea of migrating VMs in a federated cloud environment to reduce carbon footprint. They move the VMs from a high carbon footprint source to a data center with access to solar energy, using a two-step approach.

Our work is different from the discussed studies, since our objective is to minimize the cost associated with both energy consumption and carbon footprint. We consider carbon cost as a function of carbon intensity and carbon tax. Moreover, regarding the energy cost, we consider overhead energy of the data center along with the energy consumed by the servers. For this purpose, we exploit a data center's PUE model as a dynamic function of IT load and outside temperature. Finally, we present efficient and dynamic two-level VM placement approaches. These approaches observe the effect of different parameters on the total green and brown energy consumption, carbon footprint, and their associated cost for the cloud provider with distributed data center sites. In addition to this, the discussed VM placement approaches consider hourly changes in outside temperature, solar energy, and variable energy pricing.

3 System Model

In this section, we first present the system architecture, its components, and their role in a cloud computing environment. Then, we will present details on the parameters that affect cloud provider's decision in placing the VM request considering energy consumption, carbon footprint, and their associated cost. Finally, we will present the objective function and relevant constraints of the model. The list of all the symbols used in this paper are given in Table 1.

The targeted system in this study is an IaaS cloud provider offering VM resources to its clients similar to Elastic Compute Cloud (EC2) service by Amazon Web Services [34]. As shown in Fig. 1, the cloud provider consists of several geographically distributed data centers connected through a carrier network. The main parties involved in a cloud computing system are the cloud provider, cloud broker and cloud users, whose roles are discussed in the following section.

3.1 System Components

3.1.1 Cloud Provider

The cloud provider consists of n data center sites, shown as a set $\mathcal{D} = \{d_1, d_2, ..., d_n\}$, distributed in different geographical locations. Each data center site, d, is connected to a backbone network to serve cloud users and uses one or more energy sources to provide electricity for its servers, networking equipment, power systems, and other devices. A data center can just use the electricity from the off-site utility grid, O, or have its own on-site or local green sources (renewable energy), G, such as wind and solar. Moreover, data centers have their local brown energy (e.g., a diesel generator), B, in case of emergencies and outages when both grid and renewable energy are not available. Data center energy sources are shown as the set $\mathcal{E} = \{G, B, O\}$. Moreover, each data center has a set of m servers, $\mathcal{S} = \{s_1, s_2, ..., s_m\}$, with different physical configurations.

3.1.2 Cloud Broker

A cloud broker is the user-facing side of the cloud provider. It receives users VM requests that need to be routed to a data center site and then be placed on a server. The cloud broker should route requests to data centers in such a way that the energy consumption, carbon footprint, and their total cost for running the incoming workload are minimized. As stated in our previous work [35], the cloud broker uses the information sent from the data center sites to the Energy and Carbon-Efficient Cloud Information Service (ECE-CIS) to perform the VM placement.

TABLE 1
Description of Symbols

Symbol	Description	Symbol	Description
$\mathcal{D}$	Set of data center sites	$\mathcal{S}$	Set of servers in a data center
$\mathcal{E}$	Set of energy sources	$\mathcal{VM}$	Set of VM requests
x_{ij}	Matrix X's element to show VMs to data centers mapping	$y_v^B/y_v^G/y_v^O$	Element v of row vector $Y^B/Y^G/Y^O$, that shows VM v mapping to local brown/local green/off-site grid energy source
z_{vm}	Matrix Z's element to show VM to servers mapping	v^L	VM v holding time
C_T	Total cost of the energy and carbon	$C(v_{ij})$	Cost of running VM i at data center j
C_E	Cost of the energy	C_F	Cost of the carbon footprint
$C_s(v)$	Cost of the server energy to run the VM v	$C_o(v)$	Cost of the overhead energy to run the VM v
$C_E(v)$	Cost of the energy to run the VM v	$E_T(v)$	Total energy to run the VM v
$E_s(v)$	Server energy to run the VM v	$E_o(v)$	Overhead energy to run the VM v
$E_s^B(v)/E_s^G(v)/\ E_s^O(v)$	Consumed local brown/ local green/off-site grid energy to run the VM v on server s	$C_E^B/C_E^G/C_E^O$	Price of the local brown/local green/ off-site grid energy
$P_{s,\text{Peak}}$	Server power consumption at peak state	$P_{s,\text{Idle}}$	Server power consumption at idle state
U_{st}	Utilization of server s at time t	$P_s^{U_{st}}$	Server power at time t and utilization U_{st}
$P_s^{U_{s(t+)}}$	Server power consumption by running the new VM and the new utilization	P_o	Overhead power
U_t	Data center utilization at time t	H_t	Data center outside temperature at time t
$E_o^B(v)/E_o^G(v)/\ E_o^O(v)$	Consumed overhead local brown/ local green/off-site grid energy to run the VM v	$C_E^B(v)/C_E^G(v)/\ C_E^O(v)$	Cost of the consumed local brown/local green/off-site grid energy to run VM v
$R_E^B/R_E^G/R_E^O$	Carbon footprint rate of local brown/local green/off-site grid energy source	$T_F^B/T_F^G/T_F^O$	Carbon tax of local brown/local green/off-site grid energy source
v^P	VM v required processing unit	v^M	VM v required memory
s^P	Server s total processing unit	s^M	Server s total memory

3.1.3 *Cloud Users*

Cloud users submit their VM requests to the cloud broker. A submitted VM request from user i, at time t can be shown as the pair $v_i = (Type, HoldTime)$. VM type is inspired by Amazon EC2 VM instance types [34] and VM hold time depends on the application that will be run on that VM. In practice, the arrival time, type and hold time of a VM is not known by the cloud provider in advance. In our model, we serve all the VMs based on their arrival time on a first-come first-serve basis. Cloud users need to have a quality of experience (QoE) that must be satisfied by the cloud provider. The QoE for the users is defined in terms of acceptance of the submitted VM requests, which means lower rejected number of VMs higher QoE for the users.

3.2 System Parameters

Before discussing the system objective and constraints, we first introduce all the system parameters that affect the power consumption, carbon footprint, and their relevant cost.

3.2.1 *Data Center Power Efficiency*

A data center's power efficiency depends on its PUE, which is a metric to quantify the overhead power, e.g., power supplies, cooling, lightning, and UPS, in support of the incoming IT load to the system. According to Rasmussen [36] and Goiri et al. [28], the PUE is dependent on the data center utilization (IT load) and outside temperature. Therefore, we model PUE as $PUE = f(ITLoad, OutsideTemperature)$.

According to Rasmuseen [36], the most important parameter that affects PUE is the load of the data center and it has a linear relation with outside temperature. They showed a data center's PUE in two graphs, first by changing the IT load (at a constant temperature) and then by the change in the outside temperature (at a constant IT load). By using those two graphs, we interpolate a hyperbola relation between PUE and

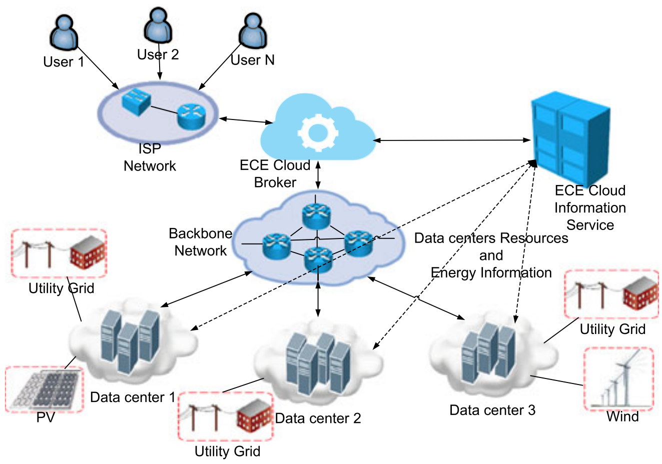

Fig. 1. System model for geographically distributed green cloud computing environment.

IT load[1] and a linear relation between PUE and outside temperature. Based on the calculations in Appendix A, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TSUSC.2017.2709980, we get

$$PUE(U_t, H_t) \simeq 1 + \frac{0.2 + 0.01U_t + 0.01U_t H_t}{U_t}. \quad (1)$$

3.2.2 Server Power Model

Each server is capable of hosting a different number of virtual machines depending on its configuration and VMs' sizes. Based on the scheduling policy, the incoming load to each server differs over time and this incoming load determines the power consumption of that server [37]. The relationship between the server power consumption and CPU utilization can be a constant, cubic, or even quadratic [38]. Attempts to make servers energy-efficient aim to make them energy proportional; which means that servers should only consume power in the presence of load [17]. A contemporary server's idle power, $P_{s,\text{Idle}}$, is half of the peak power, $P_{s,\text{Peak}}$. In this work, we use SpecPower benchmark [39] measurements to depict the relationship between server power consumption and server utilization. According to this data, a server's total drawn power increases linearly with the increase in utilization. This means that we let server's utilization be a direct mapping of CPU utilization, U_{st}. A server's power consumption as a function of CPU utilization is modeled as

$$P_s^{U_{st}} = P_{s,\text{Idle}} + (P_{s,\text{Peak}} - P_{s,\text{Idle}})U_{st}. \quad (2)$$

3.2.3 Renewable Energy Sources

Large cloud providers use renewable energy to reduce their dependency on the electricity delivered from the grid as it is costly and less clean [40], [41]. The global amount of electricity derived from renewable sources doubled between 2000 and 2012 [42] and amongst these renewable energy, wind and solar photovoltaics (PV) are the fastest growing ones. Many cloud providers try to partially get their power from renewable energy and have their own on-site solar panels and wind turbines (e.g., Facebook [43], Apple [44], and Green House Data [45]).

Most sources of renewable energy are intermittent meaning that their availability changes uncontrollably and unpredictably over time. Cloud providers can benefit from the difference of renewable energy sources at different data center sites with different time zones at the time of VM scheduling. Several studies consider how to schedule incoming workload to manage the intermittent renewable energy. Some works use the immediate available renewable energy and cancel the running jobs when the amount of solar or wind is too low or they are not available in the system [46]. Other studies consider using prediction models for the availability of this energy to assign the workload when this energy is available and reduce the job cancellation [47]. Adding storage to the data centers, where they can store the renewable energy and use it constantly in the system, is another way to overcome the unpredictability of wind and solar [48]. However, this approach has many problems [49]. For example, 1) batteries incur energy losses due to internal resistance and self-discharge, 2) battery-related costs can dominate the cost of renewable power systems, and 3) batteries use chemicals that are harmful to the environment. Given these problems, the best way to take full advantage of the available green energy is to match the energy demand to the energy supply and maximize renewable energy utilization.

In this paper, we consider solar energy as the local renewable energy for data center sites. We only take into account day/night differences for this energy. Moreover, we consider that renewable energy has the highest priority amongst all other energy sources and data centers get their power from these sources as long as they are available to have the highest renewable energy utilization.

3.2.4 Energy Price

The major incremental electricity cost of a data center is determined by the amount of energy purchased from the off-site utility grid providers. Since renewable energy has a fixed installation cost and maintenance during time, the incremental cost for using them when they are available is negligible. Moreover, the on-site brown energy (e.g., diesel generators) is only used in the absence of other energy sources. Note that we consider on-site brown energy as part of the model for the sake of comprehensiveness. However, we do not explore its effect in the evaluation part of this paper and leave it for the interested readers to simply consider it as part of their evaluation. For the electricity driven from the grid, we consider variable energy pricing during times of the day, as having on-peak and off-peak prices. By this approach, having geographically distributed data centers for a cloud provider and variable energy pricing, gives the provider the opportunity to route requests to the data center with lowest energy price. We use C_E^O, C_E^G, and C_E^B for off-site utility grid, on-site green, and on-site brown energy prices respectively, based on cents per kilowatt-hour energy usage (cents/kWh).

3.2.5 Carbon Footprint Rate and Carbon Tax

Depending on the source of the power, carbon intensity could vary significantly. We represent the carbon intensity of the energy sources by R_E^O, R_E^G, and R_E^B for off-site grid, local green, and local brown, respectively based on tons per megawatt-hour used electricity (Tons/MWh). The carbon intensity for green energy (solar and wind) is zero but brown energy, from polluting energy sources, could have different rates depending on the type of the fuel burnt to generate the electricity. As green energy availability varies during the day, one data center could get the off-site grid power from more than one provider with different carbon intensities. Moreover, to reduce the effect of the emitted CO_2 and the green house gases (GHG) on the climate change [50], carbon taxes are levied. We represent carbon tax as T_F^O, T_F^B, and T_F^G for off-site utility grid, on-site brown, and on-site green energy, respectively, as dollar per ton of the emitted carbon footprint (Dollar/Ton). We should note that the carbon intensity and carbon tax for the renewable energy are zero.

3.3 System Objective Function and Constraints

In this section, we study the objective function of the proposed system model and its constraints.

1. In this paper, we use IT load and utilization interchangeably.

3.3.1 Objective Function and Cost Modeling

The objective function is to minimize the cost of running the workload in the system, based on energy consumption and carbon footprint for the cloud provider. Meanwhile, the cloud provider should meet the cloud users' expected QoE.

The cost of running the workload is

$$C_T = \sum_{i \in \mathcal{VM}} \sum_{j \in \mathcal{D}} C(v_{ij}) x_{ij}, \tag{3}$$

where x_{ij} is an element of the two-dimensional matrix X and shows VM assignment to the data center site. If the element in this matrix is set to 1 means that v_i is assigned to d_j. Note that the summation is over the VM set, $\mathcal{VM} = \{v_1, v_2, \ldots, v_k\}$, rather than over time, since in each time epoch a data center can use more than one energy source. This means that at a certain time epoch at the data center, two running VMs could use two different energy sources. The total cost of running VMs on the servers located in geographically distributed data centers in (3) is composed of the cost of the energy used in the system plus the cost related to the carbon footprint in the environment due to the used electricity. We break this objective into an energy cost C_E and a carbon footprint cost C_F, as

$$C_T = C_E + C_F. \tag{4}$$

The energy and carbon footprint costs calculation is explained as follows.

Energy Cost. The energy cost, C_E, is the total amount of money paid to the grid electricity providers, excluding any carbon tax. In order to compute the total electricity draw in the data center sites, we need to compute the total energy used by the IT devices plus the overhead energy to run each VM. The major component of the consumed energy by the IT devices is the energy used by the servers. Therefore, we use the servers energy consumption for each VM as the total energy used by IT devices. Based on this, we can formulate the cost for the energy consumption as

$$C_E = \sum_{v \in \mathcal{VM}} (C_s(v) + C_o(v)). \tag{5}$$

Depending on the type of the energy used by the server in a data center, the cost for the energy consumption by that server is different. As mentioned earlier, a server could get its energy from three different sources: local brown, local green, and off-site brown. By having three different types of energy sources, we can formulate the cost of server energy consumption as

$$C_s(v) = \sum_{\tau \in \{B,G,O\}} E_s^\tau(v) C_E^\tau. \tag{6}$$

The energy consumption for each VM, $E_s^\tau(v)$, based on the energy source is

$$\begin{aligned} E_s^B(v) &= y_v^B E_s(v) \\ E_s^G(v) &= y_v^G E_s(v) \\ E_s^O(v) &= y_v^O E_s(v). \end{aligned} \tag{7}$$

Here, elements y_v^B, y_v^G, and y_v^O belong to row vectors Y^B, Y^G, and Y^O, respectively. If the element y_v^τ of row vector Y^τ is set to 1, means VM v is assigned to that energy source. In order to compute the energy used by each server, we compute the increase in the power consumption due to running the new VM times its holding time. The increase in energy by using server s's $\triangle P_s$ is

$$E_s(v) = \triangle P_s v^L \tag{8}$$

where the increase in power consumption,

$$\triangle P_s = (P_s^{U_{s(t+)}} - P_s^{U_{st}}), \tag{9}$$

is based on the increase of the server utilization in the next time epoch $(t+)$; that is after the VM has entered service. Based on (9) and using the server power model (2), we have

$$\triangle P_s = (P_{s,\text{Peak}} - P_{s,\text{Idle}})(U_{s(t+)} - U_{st}). \tag{10}$$

The second parameter of the energy cost function in (5) is the cost associated with the overhead energy consumed to run the VMs, (11). Depending on the type of the energy used (local brown, local green or off-site brown) the energy price would be different.

$$C_o(v) = \sum_{\tau \in \{B,G,O\}} E_o^\tau(v) C_E^\tau. \tag{11}$$

Similar to the energy cost by the servers, we calculate the overhead energy. As stated in (12), we use the VM to energy sources mapping matrices to specify the energy source used for overhead to run the incoming VM.

$$\begin{aligned} E_o^B(v) &= y_v^B E_o(v) \\ E_o^G(v) &= y_v^G E_o(v) \\ E_o^O(v) &= y_v^O E_o(v). \end{aligned} \tag{12}$$

To compute the overhead energy usage by the VM (13), we use the same approach used in (8) for calculation of the increase in the power consumption.

$$E_o(v) = \triangle P_o v^L. \tag{13}$$

As noted earlier, we use PUE as a metric to compute the overhead power consumption. PUE and overhead power relation is

$$\begin{aligned} PUE(U_t, H_t) &= \frac{P_{Total}}{P_s^{U_{st}}} = \frac{P_o + P_s^{U_{st}}}{P_s^{U_{st}}} \\ P_o &= P_s^{U_{st}}(PUE(U_t, H_t) - 1). \end{aligned} \tag{14}$$

By using (14), we can rewrite (13) as

$$\begin{aligned} E_o(v) &= (P_s^{U_{s(t+)}} - P_s^{U_{st}})(PUE(U_t, H_t) - 1)v^L \\ &= \triangle P_s v^L (PUE(U_t, H_t) - 1). \end{aligned} \tag{15}$$

Carbon Footprint Cost. The second term in the objective function, (4), is the cost of the carbon footprint contributed to the environment due to the energy consumption. We can formulate it as the product of the cost of the consumed energy, the carbon intensity, and the carbon tax of the relevant energy source. Thus, the carbon footprint cost is defined as

$$C_F = \sum_{v \in \mathcal{VM}} \sum_{\tau \in \{B,G,O\}} C_E^{\tau}(v) R_E^{\tau} T_F^{\tau}. \tag{16}$$

By using the row vectors of energy sources to VM requests mapping, we have

$$\begin{aligned} C_E^B(v) &= y_v^B C_E(v) \\ C_E^G(v) &= y_v^G C_E(v) \\ C_E^O(v) &= y_v^O C_E(v). \end{aligned} \tag{17}$$

As carbon intensity and carbon tax are zero for renewable energy sources and on-site brown is just used in the absence of the other two energy sources, we can rewrite (16) as

$$C_F = \sum_{v \in \mathcal{VM}} C_E^O(v) R_E^O T_F^O. \tag{18}$$

3.3.2 Constraints

The objective function $minimize\ C_T = C_E + C_F$ is subject to the following constraints:

- The total allocated capacity to the VM requests running on a server should not exceed the server's capacity in terms of processing unit and memory usage:

$$\begin{aligned} \sum_{v \in \mathcal{VM}} \sum_{m \in \mathcal{S}} v^P z_{vm} &\leq s^P, \\ \sum_{v \in \mathcal{VM}} \sum_{m \in \mathcal{S}} v^M z_{vm} &\leq s^M, \end{aligned} \tag{19}$$

  where, z_{vm} is an element of the two-dimensional matrix Z that is 1 if VM v is hosted on server m and 0 otherwise.
- Each running VM on a server should just use one energy source at each time epoch:

$$\forall v \in \mathcal{VM}, y_v^B + y_v^G + y_v^O = 1. \tag{20}$$

- Each element of the assigned energy sources to the VMs matrices should be greater or equal to zero:

$$y_v^B, y_v^G, y_v^O \geq 0. \tag{21}$$

- The total amount of local green energy and local brown energy consumed by VMs should not exceed the total available green and brown energy at each data center, respectively:

$$\begin{aligned} \sum_{v \in \mathcal{VM}} (E_s^G(v) + E_o^G(v)) &\leq Total\ Available\ G, \\ \sum_{v \in \mathcal{VM}} (E_s^B(v) + E_o^B(v)) &\leq Total\ Available\ B. \end{aligned} \tag{22}$$

- The total consumed off-site grid energy should not go beyond what the cloud provider receives from the electricity provider:

$$\sum_{v \in \mathcal{VM}} (E_s^O(v) + E_o^O(v)) \leq Total\ Assigned\, O. \tag{23}$$

With the definitions in Section 3.3.1, the optimization problem becomes

$$\begin{aligned} \min_{x,y} \quad & C_T \\ \text{s.t.} \quad & (19) - (23) \end{aligned} \tag{24}$$

In addition to the hard constraints, we want to give local green energy the highest priority. If there is not enough green energy available, the cloud provider uses off-site grid energy; otherwise it should use the local brown energy stored in the data center sites. That is,

$$\text{Priority}\, E^G > \text{Priority}\, E^O > \text{Priority}\, E^B.$$

4 VM Placement Approaches

In this section, we propose a dynamic VM placement algorithm to approximate (24) and six variations that neglect different components of the cost, to study the effect of different parameters and combinations of them on the amount of green and brown energy usage, carbon footprint, and total energy and carbon cost of the cloud data centers.

4.1 Cost and Renewable-Aware with Dynamic PUE (CRA-DP)

Upon the arrival of each VM request, the cloud broker has several choices with multiple data center sites and several servers within each data center, to perform VM placement. We see VM placement as a bin-packing problem with different bin sizes (e.g., physical servers) in terms of: energy price, carbon intensity, carbon tax, outside temperature, available green energy, and data center load. These differences can affect the overall energy consumption, carbon footprint, and their associated cost. Since the nature of a bin-packing problem is NP-hard, the first algorithm we propose is a derivative of the best-fit heuristic.

The CRA-DP algorithm, like that of [35], first selects the data center and then selects the server within the data center. It selects the data center with the minimum added cost for the cloud provider (minimum $\triangle C_T$), considering available green energy and dynamic PUE. CRA-DP sorts the data center sites in increasing order of the added cost due to the energy consumption and carbon footprint to run the VM for its lifetime. The server selection policy for all the algorithms in this paper is based on the least increase in the server power consumption, given by (9). The pseudocode of the algorithm is presented in Algorithm 1. Note that we do not write the rest of the algorithms pseudocode, since they all are derived from CRA-DP.

4.2 Cost-Aware with Dynamic PUE (CA-DP)

The CA-DP algorithm differs from CRA-DP in that CA-DP does not consider the availability of renewable energy while calculating the $\triangle C_T$ to select the data center site. Note that all the algorithms assume that if a data center site has renewable energy available, all the servers and racks are always powered by green energy, unless there is not enough renewable energy in the system. In this case, they will get their required power from off-site grid energy sources. The pseudocode of this algorithm is the same as the CRA-DP, but it omits Lines 8-15 and at Line 16, the $usedGreen$ is set to 0.

Algorithm 1. Cost and Renewable-Aware with Dynamic PUE (CRA-DP) VM Placement Algorithm

Input: $datacenerList$, $hostList$
Output: $destination$
1: **while** $vm\ Request$ **do**
2: Get data centers' Information from ECE-CIS;
3: **foreach** $data\ center$ in $data\ centerList$ **do**
4: $avgVmUtil \leftarrow v^P / avg\, s^P$;
5: $\triangle E_s(v) \leftarrow v^L \times P_s^{avgVmUtil}$;
6: $\triangle E_o(v) \leftarrow \triangle E_s(v) \times PUE(U_{t+}, H_t)$;
7: $\triangle E_T(v) \leftarrow \triangle E_s(v) + \triangle E_o(v)$;
8: $availGreen \leftarrow$ Get Current $availablegreenenergy$;
9: **if** $availGreen; > 0$ **then**
10: **if** $availGreen; <= \triangle E_T(v)$ **then**
11: $usedGreen \leftarrow availGreen$;
12: $availGreen \leftarrow 0$;
13: **else**
14: $usedGreen \leftarrow \triangle E_T(v)$;
15: $availGreen \leftarrow availGreen - usedGreen$;
16: $usedOffSiteEnergy \leftarrow \triangle E_T(v) - usedGreen$;
17: $\triangle C_E \leftarrow usedOffsiteEnergy \times C_E^O$;
18: $\triangle C_F \leftarrow usedOffsiteEnergy \times R_E^O \times T_F^O$;
19: $\triangle C_T \leftarrow \triangle C_E + \triangle C_F$;
20: Add $dataCenter$ with $\triangle C_T$ into $aggregateDCList$;
21: Sort $aggregateDCList$ in an ascending order of $\triangle C_T$;
22: **foreach** $dataCenter$ in $aggregateDCList$ **do**
23: **foreach** $host$ in $hostList$ **do**
24: $P_s^{U_{st}} \leftarrow$ Get current $hostDynamicPower$;
25: $P_s^{U_{s(t+)}} \leftarrow$ Calculate $hostDynamicPower$ after initiating the vm;
26: $\triangle P_s \leftarrow P_s^{U_{s(t+)}} - P_s^{U_{st}}$;
27: Sort $hostList$ in an ascending order of $\triangle P_s$;
28: **foreach** $host$ in $hostList$ **do**
29: **if** $host$ is suitable for vm **then**
30: $destination \leftarrow (data\ center, host)$;
31: return $destination$;
32: $destination \leftarrow null$; //rejection of request;
33: return $destination$;

4.3 Energy and Renewable-Aware with Dynamic PUE (ERA-DP)

The ERA-DP algorithm makes decision based on the increase in the total energy consumption (server energy + overhead energy). It calculates the total energy added to each data center to run the new VM ($\triangle E_T(v) = \triangle E_s(v) + \triangle E_o(v)$) with considering dynamic PUE and amount of the available renewable energy. This algorithm omits Lines 17-19 of the Algorithm 1 and Lines 20 and 21 are based on the $usedOffsiteEnergy$ instead of $\triangle C_T$. The rest of the algorithm is the same as the CRA-DP algorithm.

4.4 Energy-Aware with Dynamic PUE (EA-DP)

The EA-DP algorithm is similar to ERA-DP, except after calculating $\triangle E_T(v) = \triangle E_s(v) + \triangle E_o(v)$ for each data center site, it does not consider the availability of renewable energy ($usedGreen$ is set to 0).

4.5 Energy-Aware with Constant PUE (EA-CP)

This algorithm is a derivation of the EA-DP, except that PUE value does not vary by the change in IT load and outside temperature and it has a constant value. In order to obtain a reasonable constant value for PUE, we calculate its average while performing the CRA-DP algorithm from a low load until data centers get fully utilized. Note that, as considering static value for PUE just multiplies the servers energy consumption by a constant value $\triangle E_T(v) = \triangle E_s(v)(1 + PUE)$, the results are expected to be the same as when the VM placement is without considering the overhead power and just based on the servers power consumption.

4.6 Carbon Footprint-Aware with Dynamic PUE (FA-DP)

This algorithm is a derivation of the ECE algorithm in our previous work [35], which considers the effect of PUE and carbon intensity while here PUE has a dynamic value. It selects the data center with the minimum value of $R_E^\tau \times PUE(U_{t+}, H_t)$ and $\tau \in \{B, G, O\}$.

4.7 Energy Price-Aware (EPA)

The energy price-aware (EPA) VM placement algorithm, upon the arrival of each VM request selects the data center site with the cheapest energy price (minimum C_E^τ and $\tau \in \{B, G, O\}$). Since green energy cost is zero, the data center site with the available green energy has the highest priority.

5 Performance Evaluation

We evaluate the performance of the proposed approaches to investigate the effect of different parameters on the total cost, brown and green energy consumption, and carbon footprint. Note that all algorithms are evaluated based on the total cost C_T described in Section 3.3.1, even though some algorithms ignore some components of the cost.

5.1 Experiment Setup

In order to evaluate the proposed approaches, we target an IaaS cloud computing environment. Since it is difficult to perform large-scale and repeatable evaluation on real infrastructures, we use simulation to conduct our experiments. The CloudSim toolkit [51] is a simulation platform that allows evaluation of virtualized cloud environments. As the core framework of CloudSim does not support energy and carbon-efficient simulation, we use the extended version developed in our earlier work [35] that enables these features. Apart from adding the energy and carbon-awareness to the CloudSim core, we add other features such as costs of the consumed energy and emitted carbon, access to renewable energy (solar energy in this paper), overhead power consumption, and dynamic PUE.

5.1.1 Data Centers Configuration

We consider four data center sites located in four cities chosen from different states in the United States at three different time zones. These cities are chosen from the Data centers Map website [52] and they are: Dallas in Texas, Richmond in Virginia, San Jose in California, and Portland in Oregon. Since they are connected to one network backbone, the number of hops a packet traverses from source to destination is between 12 and 14 hops [53]. Therefore, different network distances do not affect site selection. Each

TABLE 2
Data Center Site Characteristics

Site Characteristics	Dallas	Richmond	San Jose	Portland
Server Power Model		$P_s^{U_t} = 120 + 154U_{st}$		
PUE Model		$PUE(U_t, H_t) = 1 + \frac{0.2+0.01U_t+0.01U_tH_t}{U_t}$		
Carbon Intensity (Tons/MWh)	0.730	0.69	0.35	0.147
Carbon Tax (Dollars/Ton)	24	22	11	48
Energy Price (cents/kWh)	6.1	6.54	10	5.77

data center has 130 heterogeneous physical servers with five different configurations described by four parameters: (Number of Cores, Core Speed (GHz), Memory (GB), Storage (GB)). The five different server types are: Type1 (2, 1.7, 16, 2000), Type2 (4, 1.7, 32, 6000), Type3 (8, 1.7, 32, 7000), Type4 (8, 2.4, 64, 7000), and Type5 (8, 2.4, 128, 9000).

5.1.2 Servers Power Consumption

As discussed in Section 3.2.2, we use the approximate linear relation with the server utilization, as shown in the work by Pellet et al. [38], for the server power model. The power model, stated in Table 2, is the linear approximation against SpecPower results for two Dell PowerEdge servers.

5.1.3 PUE Model

We use the PUE model described in Section 3.2.1 for all the data centers. We assume that the efficiencies of all the data center sites' infrastructure is the same. The PUE model is shown in Table 2.

5.1.4 Solar Energy

We use the data reported in the project undertaken by the NREL [54] to get the solar energy availability in the four aforementioned cities. We use the data of a primary station, solar radiation for flat-plate collectors facing south at a fixed tilt in ($kWh/m^2/day$). We consider five days form May 26th, 2014 to May 30th, 2014 for the simulation time and set the total area for the solar irradiation absorber flat-plates $2684m^2$ from the configuration by Solarbayer [55]. With this information, we can get the daily solar energy in terms of kWh/day. To get the hourly solar traces, we assume that the solar energy for times before 6 a.m. and after 6 p.m. is 0. Moreover, the distribution of the energy between 6 a.m. and 6 p.m. has a raised cosine distribution, with the peak at 12 noon. Knowing the total solar of one day and integrating the raised cosine between 6 a.m. and 6 p.m., we calculate hourly available solar energy in kWh for these four cities as shown in Fig. 2.

5.1.5 Carbon Footprint Rate and Carbon Tax

The data centers' carbon intensity ($Tons/MWh$) is obtained from the US Department of Energy, Appendix F, available in the online supplemental material, Electricity Emission Factors [56]. We use the data reported by the Carbon Tax Center [57] for the carbon tax, due to the contribution in emitting carbon in the environment, in terms of dollars per ton of CO_2 ($Dollars/Ton$). Values for carbon intensity and carbon tax for the chosen data center sites are reported in Table 2.

5.1.6 Energy Price

We consider on-peak and off-peak pricing model for the electricity driven from off-site electricity providers. Energy prices are taken from the US Energy Information Administration [58]. Peak energy price for 4 sites are shown in Table 2. Times of the day before 8 a.m. and after 10 p.m. are off-peak times and the energy price will be half of the on-peak times (8 a.m. to 10 p.m.). We assume the on-site solar energy has zero incremental cost, since it has a one time capital cost and regular maintenance independent of use.

5.1.7 Outside Temperature

We derive the hourly temperature of the four data center sites from May 26th, to May 30th 2014 from the Weatherbase portal [59]. Fig. 3 shows the hourly temperature for the aforementioned sites.

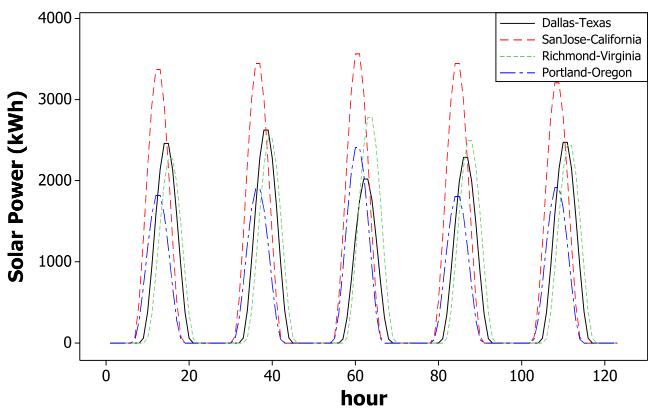

Fig. 2. Solar energy for five days.

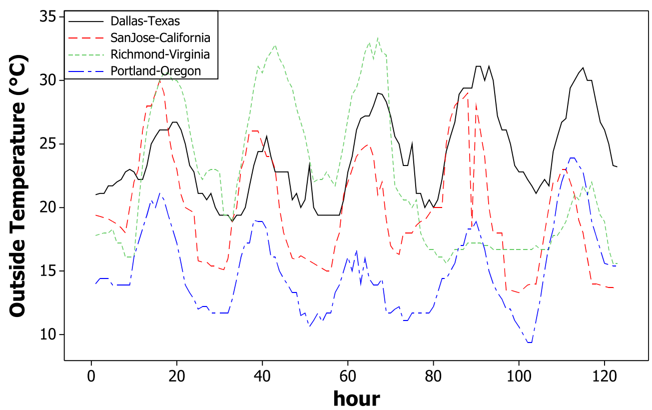

Fig. 3. Outside temperature for five days.

TABLE 3
VM Types and Simulated User Requests (Bag-of-Task (BT) and Web-Request (WR))

VM Type		Number of Cores	Core Speed (GHz)	Memory (MB)	Storage (GB)	Probability and UserRequest
Standard Instances	M1Small	1	1	1740	160	0.25-BT
	M1Large	2	4	7680	850	0.12-WR 0.25-BT
	M1XLarge	4	8	15360	1690	0.08-WR
High Memory Instances	M2XLarge	2	6.5	17510	420	0.12-WR
	M22XLarge	4	13	35020	850	0.08-WR
High CPUInstances	C1Medium	2	5	1740	320	0.1-BT

5.1.8 Workload Data

The incoming workload to the system is the VM requests from cloud users. Since we only deal with the placement of the VM requests and allocation of their required resources, we do not need to know the type of application running within the instantiated VM. However, we assume that each VM operates at its maximum utilization and uses all the allocated resource. Each VM request has physical characteristics, that are inspired by Amazon EC2 VM instance types. Beside the physical requirements, each VM has a submission time from the user and holding time. For the system workload, we use the same model and workload generator we used before [35]. We generate two types of VMs known as bag-of-tasks and web-requests with the same arrival rate and different holding time pattern (longer holding time for web-requests). The applied workload generator for this purpose is the Lublin-Feitelson [60] workload model. To generate bag-of-tasks, we use the parameters from [60], except that we change the first parameter of the Gamma distribution to 20.4 to get VMs with longer holding times, and we change the holding-time distribution to Hyper Gamma, with mean 73 and variance 165, to generate web-requests. The VM types and the probability of each type submitted from the users are stated in Table 3.

We ran the simulation for 5 days (120 hours) and in order to have a steady environment, we omitted 5 percent of the generated requests from the start and 5 percent from the end as they are part of the warm-up and cool-down of the system, respectively. (The latter is necessary as the CloudSim simulation finishes when the last VM completes.) Note that we consider each request generated by Lublin as a VM request. Finally, since Lublin takes a random number as input, we repeated each experiment 30 times, and report the mean of the results.

5.2 Experiment Results

In the experiments, we measure the total amounts of green and brown energy consumption, carbon footprint, and their associated cost. Moreover, we check the total cost of the cloud computing system under different VM placement policies. Finally, we measure the number of rejected VMs in the system due to insufficient physical resources that leads to the violation of users' QoE in terms of SLA violation. The load varied from 500 VMs, to show how the system behaves when one data center has the physical capacity to host all the requests, up to 1700 VMs, when the system performs at its full utilization and rejects some of the incoming load.

Note that in the experiments, we checked that the results are not skewed and based on this we report their general behavior on the mean value. Moreover, we performed 2-sample t-test to check whether the differences in results are significant or not.

5.2.1 Green Energy Consumption

In this experiment, we measure the amount of green energy consumed by different VM placement policies to run the incoming workload in the system. As Fig. 4 demonstrates, three algorithms (ERA-DP, CRA-DP, EPA) that consider availability of renewable energy in the placement, have the most green energy consumption, with a slightly higher usage for the ERA-DP algorithm. The EA-CP algorithm has the smallest green energy consumption, as it is not renewable-aware and uses a constant value for PUE. The latter factor leads to not considering data centers' load change and their outside temperature; therefore it does not lead to an efficient site selection and distribution of load among data centers to get the most of available solar energy at different times of the day. In order to study the effect of considering dynamic PUE versus constant PUE and renewable energy, we run a 2-sample t-test on ERA-DP and EA-CP. We get $p = 0.04$, therefore we conclude that considering dynamic PUE and renewable energy have significant effect on the total green energy consumption. The algorithms (CA-DP, EA-DP, and FA-DP) that are not renewable-aware consume less green energy as well. But the difference with the group of renewable-aware algorithms is not significant ($p > 0.05$), since, as noted earlier, green energy has the highest priority if the data center has access to it.

5.2.2 Brown Energy Consumption

Fig. 5 shows the amount of brown energy consumption by different VM placement policies. At lower loads, EA-CP

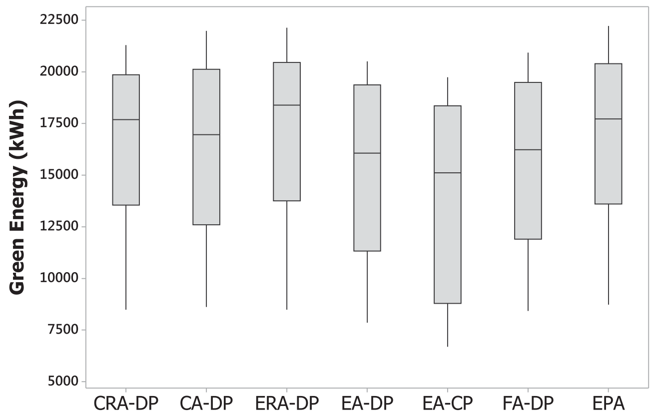

Fig. 4. Green energy consumption.

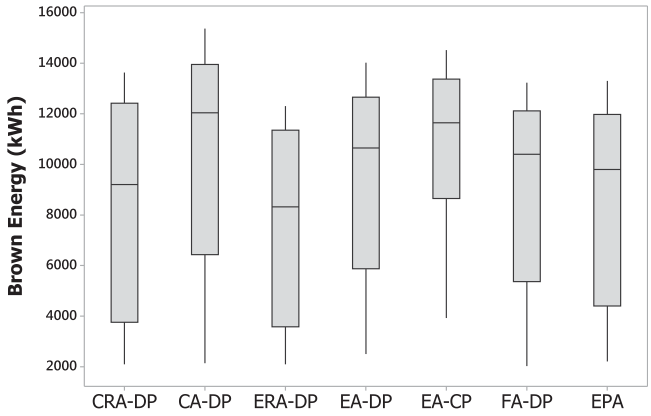

Fig. 5. Brown energy consumption.

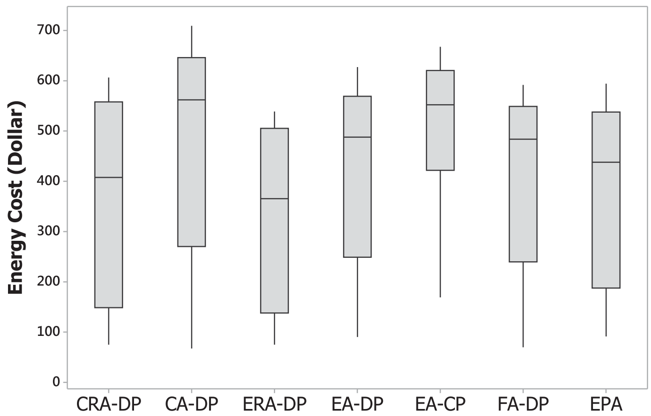

Fig. 6. Energy cost.

consumes significantly more brown energy than the other algorithms, as it is based on a constant value for PUE and distributes the load without considering current load of the data center sites and the outside temperature. The rest of the policies have close behavior. The reason is that they all are based on dynamic PUE, the system load is low and the renewable energy source has the highest priority. As the system load increases, EA-CP continues consuming more brown energy, just with a slight improvement; since the constant PUE value, that is the average value of PUE gets closer to the real dynamic value. From the results we observe that CA-DP has a sudden increase in the brown energy consumption. Because its placement is based on the increase in the total cost in the system and parameters, such as dynamic energy pricing, that affect the decision making do not have any impact on reducing the total brown consumption.

Overall, ERA-DP policy has the lowest brown energy consumption. It consumes, on average, 8.9 percent less brown energy in comparison to its competitor, CRA-DP algorithm. These two algorithms along with EPA, that is also based on considering renewable availability, have the lowest brown energy consumption. Moreover, ERA-DP, consumes 31.3 percent less brown energy on-average than EA-CP and 36.4 percent less than CA-DP. Based on the 2-sample t-test on ERA-DP and EA-CP, there is significant difference ($p = 0.01$) in the amount of consumed brown energy. Moreover, t-test on ERA-DP and CA-DP shows the significance ($p < 0.05$) of considering increase in the energy

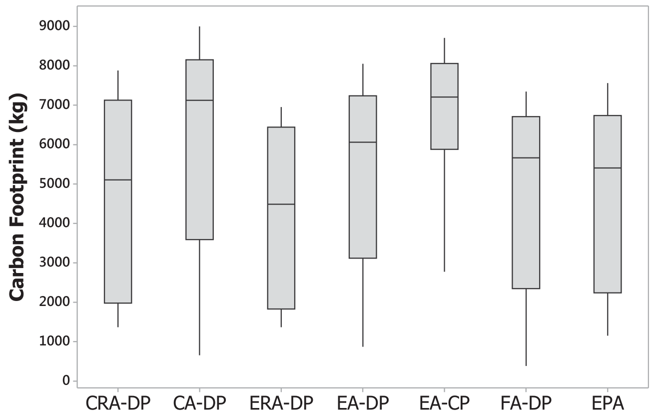

Fig. 7. Carbon footprint.

consumption rather than increase in the total cost while VM placement is carried out.

5.2.3 Energy Cost

Energy cost is a function of the amount of brown energy consumption, since the cost of renewable energy is considered zero in this paper. We observe the same behavior among the algorithms in Fig. 6 as we witnessed in Fig. 5. Algorithm ERA-DP reduces the energy cost by an average of 10.03 percent compared to its competitor algorithm, CRA-DP. Moreover, t-test results show that the energy cost difference between ERA-DP and two other algorithm, EA-CP and CA-DP, is significant with $p = 0.002$ and $p = 0.042$, respectively. This emphasizes the importance of considering dynamic PUE, renewable energy, and increase in energy consumption.

5.2.4 Carbon Footprint

Carbon footprint in the system, likewise energy cost, is a result of the usage of the brown energy sources. Hence, we should expect a similar pattern as Fig. 5. But we should not expect the same gap from one policy to another, since different energy sources have different carbon intensities. One significant difference in Fig. 7 is that, at lower workload (VM < 800), FA-DP performs significantly better than ERA-DP. The reason is that FA-DP considers sources carbon intensity and dynamic PUE at the same time and at lower loads it submits the requests to the data center with the minimum $carbon\ footprint \times PUE$. Though by the increase in the incoming load and the need to use more than one site, this policy does not perform optimal and ERA-DP is the algorithm that has a better performance. Overall, ERA-DP comparing to its close competitor, FA-DP, reduces carbon footprint 10.6 percent on average. In addition, it reduces carbon footprint on an average of 60 and 42 percent in comparison to EA-CP and CA-DP, respectively. T-test shows $p < 0.01$ and $p = 0.044$ for ERA-DP versus EA-CP and CA-DP, respectively, which again assures the importance of considering dynamic PUE, renewable energy, and changes in energy consumption.

5.2.5 Carbon Cost

Fig. 8 shows the cost of the carbon footprint in dollars. Since any increase in the value of a carbon tax is the result of

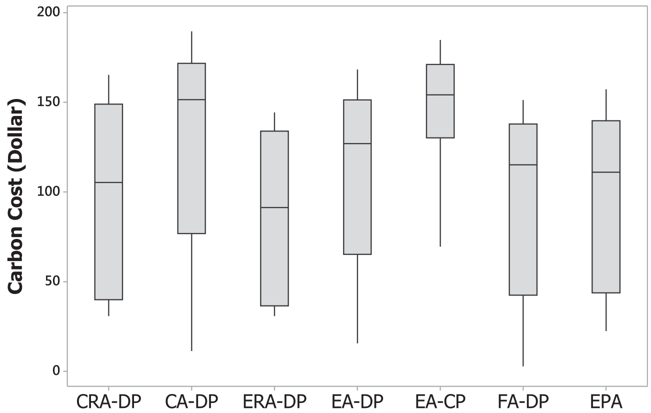

Fig. 8. Carbon cost.

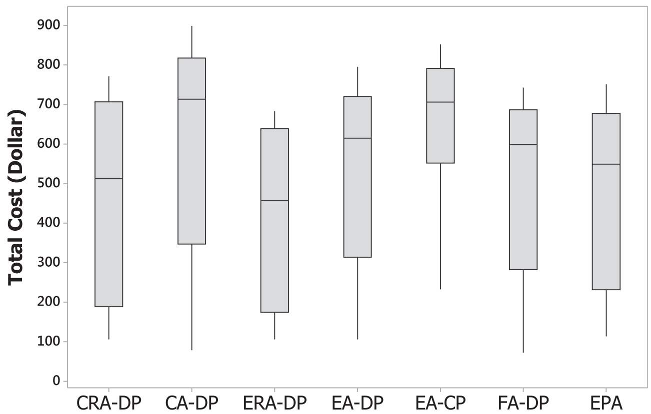

Fig. 9. Total cost.

carbon footprint growth, the behavior of different algorithms and their gaps would be the same as the total carbon footprint in Fig. 7. Still ERA-DP on average has 7.4 percent less carbon cost comparing to FA-DP. Moreover, it has on average 68 percent and 45.6 percent better performance comparing to EA-CP ($p < 0.01$) and CA-DP ($p = 0.33$), respectively.

5.2.6 Total Cost

Fig. 9 demonstrates an overall view of the effect of different VM placement policies on the total cost related to the energy and carbon footprint. At lower system loads, carbon cost (FA-DP) has a slight effect on the total cost of the system; whilst with the increase in the load, ERA-DP improves the total cost by an average of 19.3 and 10.5 percent comparing to FA-DP and CRA-DP, respectively. Moreover, ERA-DP significantly improves the total cost by an average of 57.3 and 43.8 percent in comparison to EA-CP ($p = 0.001$) and CA-DP ($p = 0.04$), respectively.

5.2.7 SLA Violation

The last experiment measures SLA violation rate in order to make sure users' quality of experience is satisfied. SLA is calculated as the number of rejected VMs due to insufficient physical resources in the system. Table 4 shows SLA violation rate under increasing workload for different VM placement policies. The table reports violations for loads from 1300, since below this load the violation rate for all the policies is zero and all the incoming load to the system are served. From the table, we observe that all the placement policies have close SLA violation. Moreover, ERA-DP at two points has the minimum violation rate and in the rest it only has 0.1-0.2 percent higher violation comparing the minimum reported ones. As a result, we can conclude that ERA-DP performs better in terms of brown energy consumption, carbon footprint, energy and carbon cost. Moreover, it has close, even at some points minimum, values for SLA violation comparing to the competitive algorithms.

TABLE 4
SLA Violation for VM Placement Policies

Algorithm	SLA Violation Under Different VM Requests				
	1300	**1400**	**1500**	**1600**	**1700**
CRA-DP	0.08%	0.3%	**0.9**%	**2.7**%	**4.9**%
CA-DP	**0.0**%	0.3%	1.1%	2.9%	5.2%
ERA-DP	**0.0**%	**0.2**%	1.0%	2.6%	5.1%
EA-DP	0.06%	0.3%	1.1%	2.9%	5.2%
EA-CP	0.05%	0.3%	1.3%	3.1%	5.5%
FA-DP	0.05%	0.3%	1.3%	3.1%	5.5%
EPA	0.14%	0.8%	2.4%	4.5%	6.3%

6 Conclusions and Future Directions

This paper investigates different parameters that affect energy and carbon cost for a cloud provider with geographically distributed data center sites. First, we consider carbon cost as part of the total cost that enables the provider not only decrease the total cost, but also reduce the CO_2 emission. Moreover, to decrease the energy cost, we consider overhead energy consumption in support of IT devices in the data center. We employ PUE as a metric that affects overhead energy of a data center, which is responsible for almost half of the energy consumption. We exploit a model for PUE as a function of data center's IT load and outside temperature. Further, we consider access to renewable energy sources, besides off-site grid (known as brown) sources.

We have presented and evaluated different energy and carbon-aware dynamic VM placement approaches. In a nutshell, ERA-DP that considers dynamic PUE, availability of renewables, and changes in energy consumption has the highest effect in reducing the total cost of energy and carbon and also reducing brown energy usage; whilst has the same level of SLA compared to the other algorithms. Furthermore, amongst the renewable-aware algorithms (CRA-DP and ERA-DP) and EPA, the later algorithm performs worse. Because EPA prefers the sites with available renewable energy, as it has the lowest price (zero), thus distributes the load between data center sites to get the most of renewables. This leads to use of computing resources of all the data centers and having overhead power as a major killer for the power consumption in all the sites. In the future, we plan to consider the effect of adding prediction of availability of renewable energy and changes in temperature on the decision making in VM placement. Moreover, we plan to provide competitive-ratio bound of the online algorithm comparing the optimal off-line for the ERA-DP algorithm. We will also consider having data center sites in different countries and the effect of carrier network delay as part of the total cost.

REFERENCES

[1] M. Armbrust, et al., "A view of cloud computing," *Commun. ACM*, vol. 53, no. 4, pp. 50–58, 2010.

[2] E. Heyd, "America's data centers consuming massive and growing amounts of electricity," 2014. [Online]. Available: https://www.nrdc.org/media/2014/140826

[3] J. Hamilton, "Overall data center costs," 2010. [Online]. Available: http://perspectives. mvdirona. com

[4] C. J. Hepburn, "Regulating by prices, quantities or both: An update and an overview," *Oxford Rev. Econ. Policy*, vol. 22, no. 2, pp. 226–247, 2006.

[5] C. Belady, A. Rawson, J. Pfleuger, and T. Cader, "Green grid data center power efficiency metrics: Pue and dcie," Technical report, Green Grid, 2008.

[6] "Latest Microsoft Data Center Design Gets Close to Unity PUE," [Online]. Available: http://www.datacenterknowledge.com/archives/2016/09/27/latest-microsoft-data-center-design-gets-close-to-unity-pue/. Accessed on: Apr. 7, 2017.

[7] A. Shehabi, et al., "United states data center energy usage report," Lawrence Berkeley Nat. Laboratory, Berkeley, CA, USA, Rep. No. LBNL-1005775, vol. 4, 2016.

[8] APAC data center survey reveals high PUE figures across the region. [Online]. Available: http://www.datacenterdynamics.com/news/apac-data-center-survey-reveals-high-pue-figures-across-the-region/75116.fullarticle. Accessed on: Apr. 7, 2017.

[9] A. Berl, et al., "Energy-efficient cloud computing," *Comput. J.*, vol. 53, no. 7, pp. 1045–1051, 2010.

[10] C.-M. Wu, R.-S. Chang, and H.-Y. Chan, "A green energy-efficient scheduling algorithm using the DVFS technique for cloud datacenters," *Future Generation Comput. Syst.*, vol. 37, pp. 141–147, 2014.

[11] T. Brey and L. Lamers, "Using virtualization to improve data center efficiency," *The Green Grid, Whitepaper*, vol. 19, 2009.

[12] S. Srikantaiah, A. Kansal, and F. Zhao, "Energy aware consolidation for cloud computing," in *Proc. Conf. Power Aware Comput. and Syst.*, 2008, pp. 10–10.

[13] E. Harney, S. Goasguen, J. Martin, M. Murphy, and M. Westall, "The efficacy of live virtual machine migrations over the internet," in *Proc. 2nd Int. Workshop Virtualization Technol. Distrib. Comput.*, 2007, pp. 1–7.

[14] C. Clark, et al., "Live migration of virtual machines," in *Proc. 2nd Conf. Symp. Netw. Syst. Des. Implementation*, 2005, vol. 2, pp. 273–286.

[15] M. Lin, A. Wierman, L. Andrew, and E. Thereska, "Dynamic right-sizing for power-proportional data centers," in *Proc. IEEE INFOCOM*, 2011, pp. 1098–1106.

[16] D. Shen, et al., "Stochastic modeling of dynamic right-sizing for energy-efficiency in cloud data centers," *Future Generation Comput. Syst.*, vol. 48, pp. 82–95, 2014.

[17] L. A. Barroso and U. Hölzle, "The case for energy-proportional computing," *IEEE Comput.*, vol. 40, no. 12, pp. 33–37, Dec. 2007.

[18] L. Lefèvre and A. Orgerie, "Designing and evaluating an energy efficient cloud," *J. Supercomputing*, vol. 51, no. 3, pp. 352–373, 2010.

[19] B. Aksanli, J. Venkatesh, L. Zhang, and T. Rosing, "Utilizing green energy prediction to schedule mixed batch and service jobs in data centers," in *Proc. 4th Workshop Power-Aware Comput. Syst.*, 2011, pp. 5:1–5:5.

[20] K. Le, R. Bianchini, M. Martonosi, and T. Nguyen, "Cost-and energy-aware load distribution across data centers," *Proc. 6th Workshop Power-Aware Comput. Syst.*, 2009, pp. 1–5.

[21] Z. Liu, M. Lin, A. Wierman, S. H. Low, and L. L. Andrew, "Geographical load balancing with renewables," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 39, no. 3, pp. 62–66, 2011.

[22] M. Lin, Z. Liu, A. Wierman, and L. L. Andrew, "Online algorithms for geographical load balancing," in *Proc. Int. Green Comput. Conf.*, 2012, pp. 1–10.

[23] C. Chen, B. He, and X. Tang, "Green-aware workload scheduling in geographically distributed data centers." in *Proc. IEEE Int. Conf. Cloud Comput. Technol. Sci.*, 2012, pp. 82–89.

[24] A. Celesti, A. Puliafito, F. Tusa, and M. Villari, "Energy sustainability in cooperating clouds," in *Proc. 3rd Int. Conf. Cloud Comput. and Serv. Sci.*, 2013, pp. 83–89.

[25] K. Le, R. Bianchini, T. D. Nguyen, O. Bilgir, and M. Martonosi, "Capping the brown energy consumption of internet services at low cost," in *Proc. Int. Green Comput. Conf.*, 2010, pp. 3–14.

[26] K. Le, R. Bianchini, J. Zhang, Y. Jaluria, J. Meng, and T. D. Nguyen, "Reducing electricity cost through virtual machine placement in high performance computing clouds," in *Proc. Int. Conf. for High Perform. Comput., Netw., Storage Anal.*, 2011, pp. 22:1–22:12.

[27] S. Ren, Y. He, and F. Xu, "Provably-efficient job scheduling for energy and fairness in geographically distributed data centers," in *Proc. 32nd Int. Conf. Distrib. Comput. Syst.*, 2012, pp. 22–31.

[28] Í. Goiri, K. Le, J. Guitart, J. Torres, and R. Bianchini, "Intelligent placement of datacenters for internet services," in *Proc. 31st Int. Conf. Distrib. Comput. Syst.*, 2011, pp. 131–142.

[29] S. K. Garg, C. S. Yeo, A. Anandasivam, and R. Buyya, "Environment-conscious scheduling of HPC applications on distributed cloud-oriented data centers," *J. Parallel Distrib. Comput.*, vol. 71, no. 6, pp. 732–749, 2011.

[30] S. K. Garg, C. S. Yeo, and R. Buyya, "Green cloud framework for improving carbon efficiency of clouds," in *Proc. 17th Int. Conf. Parallel Process.*, 2011, pp. 491–502.

[31] N. Buchbinder, N. Jain, and I. Menache, "Online job-migration for reducing the electricity bill in the cloud," in *Proc. Int. Conf. Res. Netw.*, 2011, pp. 172–185.

[32] M. Giacobbe, A. Celesti, M. Fazio, M. Villari, and A. Puliafito, "An approach to reduce energy costs through virtual machine migrations in cloud federation," in *Proc. Symp. Comput. Commun.*, 2015, pp. 782–787.

[33] M. Giacobbe, A. Celesti, M. Fazio, M. Villari, and A. Puliafito, "Evaluating a cloud federation ecosystem to reduce carbon footprint by moving computational resources," in *Proc. Symp. Comput. Commun.*, 2015, pp. 99–104.

[34] Amazon Web Services. [Online]. Available: http://aws.amazon.com/. Accessed on: Mar. 9, 2017.

[35] A. Khosravi, S. K. Garg, and R. Buyya, "Energy and carbon-efficient placement of virtual machines in distributed cloud data centers," in *Proc. Euro-Par Parallel Process.*, 2013, pp. 317–328.

[36] N. Rasmussen, "Electrical efficiency measurement for data centers," *White paper*, vol. 154, 2007.

[37] M. Blackburn and G. Grid, *Five Ways to Reduce Data Center Server Power Consumption*. The Green Grid, 2008.

[38] S. Pelley, D. Meisner, T. F. Wenisch, and J. W. VanGilder , "Understanding and abstracting total data center power," in *Proc. Workshop Energy-Efficient Des.*, 2009.

[39] "Standard performance evaluation corporation. SPECpower" 2008. [Online]. Available: http://www.spec.org/power_ssj2008

[40] J. Peterpaul, "Solar panel module and support therefor," US Patent 4 636 577, Jan. 13 1987.

[41] P. Thibodeau, "Wind power data center project planned in urban area," *ComputerWorld*, Apr. 2008.

[42] NREL: News - NREL releases renewable energy data book detailing growing industry. [Online]. Available: http://www.nrel.gov/docs/fy17osti/66591.pdf. Accessed on: Apr. 7, 2017.

[43] Facebook Installs Solar Panels at New Data Center. [Online]. Available: http://www.datacenterknowledge.com/archives/2011/04/16/facebook-install%s-solar-panels-at-new-data-center/. Accessed on: Jan. 30, 2015.

[44] Apple Plans 20MW of Solar Power for iDataCenter. [Online]. Available: http://www.datacenterknowledge.com/archives/2012/02/20/apple-plans-20mw-of-solar-power-for-idatacenter/. Accessed on: Jan. 30, 2015.

[45] Wind-Powered Data Center in Wyoming. [Online]. Available: http://www.datacenterknowledge.com/archives/2007/11/29/wind-powered-data-center-in-wyoming/. Accessed on: Jan. 30, 2015.

[46] A. Krioukov, C. Goebel, S. Alspaugh, Y. Chen, D. E. Culler, and R. H. Katz, "Integrating renewable energy using data analytics systems: Challenges and opportunities," *IEEE Data Eng. Bull.*, vol. 34, no. 1, pp. 3–11, Mar. 2011.

[47] B. Aksanli, J. Venkatesh, L. Zhang, and T. Rosing, "Utilizing green energy prediction to schedule mixed batch and service jobs in data centers," *ACM SIGOPS Operating Syst. Rev.*, vol. 45, no. 3, pp. 53–57, 2012.

[48] D. Gmach, et al., "Capacity planning and power management to exploit sustainable energy," in *Proc. Int. Conf. Netw. Service Manag.*, 2010, pp. 96–103.

[49] Í. Goiri, K. Le, T. D. Nguyen, J. Guitart, J. Torres, and R. Bianchini, "Greenhadoop: Leveraging green energy in data-processing frameworks," in *Proc. 7th ACM Eur. Conf. Comput. Syst.*, 2012, pp. 57–70.

[50] P. Hoeller and M. Wallin, *Energy Prices, Taxes and Carbon Dioxide Emissions*. Paris, France: OECD Paris, 1991.
[51] R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. De Rose, and R. Buyya, "Cloudsim: A toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms," *Softw.: Practice Experience*, vol. 41, no. 1, pp. 23–50, 2011.
[52] Data center map. [Online]. Available: http://www.datacentermap.com/. Accessed on: Jan. 30, 2015.
[53] P. Van Mieghem, *Performance Analysis of Communications Networks and Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2006.
[54] Solar Radiation Data Manual for Flat-Plate and Concentrating Collectors. [Online]. Available: http://rredc.nrel.gov/solar/pubs/redbook/. Accessed on: Jan. 30, 2015.
[55] Solarbayer. [Online]. Available: http://www.solarbayer.com/. Accessed on: Jan. 30, 2015.
[56] US Department of Energy, Appendix F, Electricity Emission Factors. [Online]. Available: http://cloud.agroclimate.org/tools/deprecated/carbonFootprint/references/Electricity_emission_factor.pdf. Accessed on: Jan. 30, 2015.
[57] Carbon Tax Center, Pricing carbon efficiently and equitably. [Online]. Available: http://www.carbontax.org/. Accessed on: Jan. 30, 2015.
[58] EIA-electricity data. [Online]. Available: http://www.eia.gov/electricity/monthly/pdf/epm.pdf. Accessed on: Apr. 7, 2017.
[59] Travel weather averages (weatherbase). [Online]. Available: http://www.weatherbase.com/. Accessed on: Jan. 30, 2015.
[60] U. Lublin and D. Feitelson, "The workload on parallel supercomputers: modeling the characteristics of rigid jobs," *J. Parallel Distrib. Comput.*, vol. 63, no. 11, pp. 1105–1122, 2003.

Atefeh Khosravi received the BSc degree in information technology engineering, in 2008, and the MSc degree in computer engineering, in 2011, both from the Amirkabir University of Technology. She is working toward the PhD degree at the Cloud Computing and Distributed Systems (CLOUDS) Laboratory within the School of Computing and Information Systems, the University of Melbourne. Her PhD research is on energy and carbon-aware resource management across distributed cloud data centers. It is mainly focused on the development of policies and algorithms for placement and migration of virtual machines by considering data centers energy sources, carbon footprint, and power consumption. She is also a research scientist with Amazon Web Services.

Lachlan L. H. Andrew (M'97-SM'05) received the BSc, BE, and PhD degrees from the University of Melbourne, Australia, in 1992, 1993, and 1997. Since 2008, he has been an associate professor at the Swinburne University of Technology, Australia, and since 2010, he has been an ARC Future fellow. From 2005 to 2008, he was a senior research engineer in the Department of Computer Science, Caltech. Prior to that, he was a senior research fellow at the University of Melbourne and a lecturer at RMIT, Australia. His research interests include energy-efficient networking and performance analysis of resource allocation algorithms. He was co-recipient of the best paper award at IGCC2012, IEEE INFOCOM 2011, and IEEE MASS 2007. He is a member of the ACM and senior member of the IEEE.

Rajkumar Buyya is a professor of computer science and software engineering, Future fellow of the Australian Research Council, and director of the Cloud Computing and Distributed Systems (CLOUDS) Laboratory with the University of Melbourne, Australia. He is also serving as the founding CEO of Manjrasoft, a spin-off company of the university, commercializing its innovations in cloud computing. He has authored more than 525 publications and seven text books including "Mastering Cloud Computing" published by McGraw Hill, China Machine Press, and Morgan Kaufmann for Indian, Chinese, and international markets, respectively. He is one of the highly cited authors in computer science and software engineering worldwide (h-index=108, g-index=230, 56500+ citations). Software technologies for grid and cloud computing developed under his leadership have gained rapid acceptance and are in use at several academic institutions and commercial enterprises in 40 countries around the world. These contributions and international research leadership are recognized through the award of the "2009 IEEE Medal for Excellence in Scalable Computin" from the IEEE Computer Society, USA. Manjrasoft's Aneka Cloud technology developed under his leadership has received the 2010 Frost & Sullivan New Product Innovation Award. He served as the foundation Editor-in-Chief (EiC) of the *IEEE Transactions on Cloud Computing* and is currently serving as Co-EiC of the *Journal of Software: Practice and Experience*. He is a fellow of the IEEE.