Carbon Footprints Estimation of a Novel Environment-Aware Cloud Architecture

Neha Solanki and Rajesh Purohit

Abstract Carbon footprints are increasing with a huge rate and the IT world is also contributing in this increase. In cloud computing, with the growth of demand for high performance computing infrastructure, number of data centers has increased. To cater the demand of high availability, the data centers are kept running round the clock. This causes high energy consumption and eventually increases in carbon footprints, which is harmful for environment. In addition to this, high energy consumption leads to costlier business. In this paper, a novel architecture for cloud is proposed by introducing an energy-aware service provider layer. The responsibility of this layer is to monitor and control the performance of cloud data centers for reducing energy consumption and carbon footprints. Live migration of virtual machines among physical machines is applied as basic technique for reducing the energy consumption.

Keywords Bin-packing algorithms · Carbon footprints · Consolidation · Data centers · Live migration

1 Introduction

Cloud computing delivers on-demand computing resources over the Internet on the basis of pay-as-you-go model. According to National Institute of Standards and Technology, USA, "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction" [1]. It provides three types of services:

Neha Solanki (✉) · Rajesh Purohit
Computer Science and Engineering Department, M.B.M Engineering College,
Jai Narain Vyas University, Jodhpur, India
e-mail: nehasolanki2789@gmail.com

Rajesh Purohit
e-mail: rajeshpurohit@jnvu.edu.in

software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS).

Virtualization is used in cloud computing which shares the underlying hardware infrastructure and through virtual machines it provides the computing resources. It gives full control to cloud provider's administrator for virtual machine allocation, which results in efficient utilization of the resources.

Cloud services are run by data centers. A single data center may consist of a large number of physical machines. These data centers consume huge amount of electricity which increases carbon footprints and the computational cost. According to National Resource Defense Council's report, energy consumed by data centers in US in 2013 was estimated as 91 billion kWh of electricity which is enough to power all the houses in New York City twice and this is annual output of 34 large (500-mW) coal-fired power plants [2]. Data center's electricity consumption is supposed to increase to around 140 billion KWh annually by 2020, which is equal to 50 power plants' annual output and emitting nearly 150 million metric tons of carbon pollution annually [2]. Hence from 2013 to 2020, data center energy consumption is projected to increase roughly 53 %.

It is necessary to reduce this high energy consumption so that carbon footprints also get reduced. The main objective of this paper is to estimate the carbon footprints of the proposed architecture for environment-aware cloud computing. CloudSim is a simulation framework developed at University of Melbourne which allows modeling/simulation and experimentation of proposed solution on specific system design issues for investigation at abstract level [3]. To convert the results of estimated energy consumption to carbon footprints, defra conversion factor is applied [4].

2 Related Work

In recent years, many researchers have invested their time on energy-saving architecture and techniques of cloud. Verma et al. [5] have developed the architecture of pMapper, a power-aware application placement controller with heterogeneous virtualized server clusters. It minimizes power and migration cost, while fulfilling the performance requirement.

Beloglazov et al. [6] developed an architectural framework and principles for energy-efficient cloud computing. This architecture aimed to improve energy efficiency of the data center, while delivering the negotiated quality of service (QoS). For this, live migration concept is used to save energy of cloud data centers. Aboozar et al. [7] have proposed decision support-as-a-service (DSaaS) architecture which is divided into two subsystems: cloud side and user side to help managers for making fast and precise decisions for energy saving. Yanfeiet et al. [8] have proposed system architecture of EECLOUD, in which the data center is divided into several time-sub-clusters to which jobs with similar runtime are assigned. It uses live migration for saving energy consumption. Jinhai et al. [9] have designed

energy-efficient architecture of cloud data center which is divided into two types of controller units: global controller and local controller. It also uses live migration to reduce energy consumption.
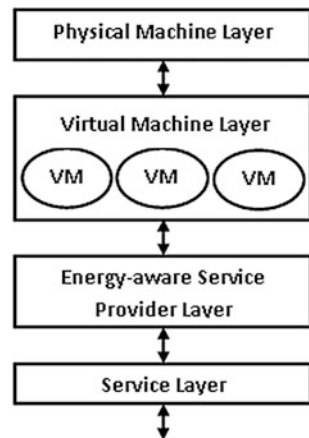
3 Environment-Aware Architectural Framework of Cloud Data Centers

The abstract view of proposed environment-aware architecture is divided into four layers, i.e., service layer, energy-aware service provider layer, virtual machine layer, and physical layer as depicted by Fig. 1.

1. Service Layer: This layer provides client interface for submitting their request through web to enable time, location, and device independency.
2. Energy-aware Service Provider Layer: This layer acts as an interface between users and cloud infrastructure and it is responsible for conserving energy of cloud data centers. It uses various live migration strategies to consolidate virtual machine to save energy [10].
3. Virtual Machine Layer: This layer consists of virtual machines (VM) which can dynamically start and stop physical machines according to incoming users' requests. On a single physical machine, multiple virtual machines can run applications concurrently.
4. Physical Machine Layer: This layer consists of computing servers which provide the hardware infrastructure like computing resource, storage resource, etc. This hardware infrastructure is used for making virtualized resources to fulfill users' service demands.

The proposed environment-aware architecture shown in Fig. 2 reduces energy consumption of cloud data centers by consolidating the allocated virtual machines

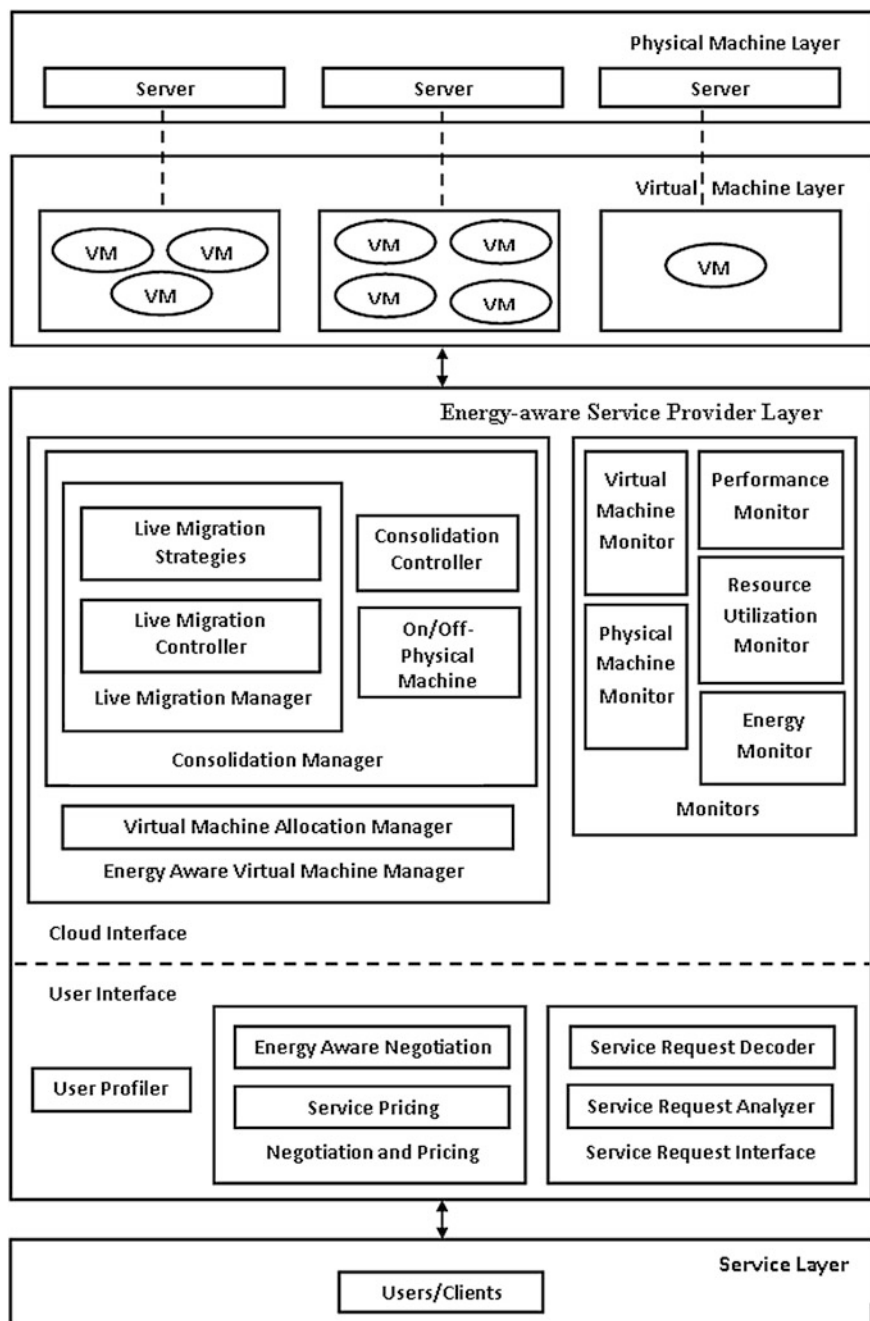**Fig. 1** Abstract view of environment-aware architecture of cloud

Fig. 2 Environment-aware architecture of cloud

to minimize the number of currently running physical machines and switching off the idle physical machines. To fulfill the requirement of new virtual machines, these switched-off physical machines are switch-on.

In the *Service Layer*, users/clients submit their service request to the cloud. *Energy-aware Service Provider Layer* which acts as an interface between users and cloud infrastructure is divided into two sub layers: *User Interface* and *Cloud Interface*.

User Interface sub layer consists of: *User Profiler*, *Negotiation and Pricing*, and *Service Request Interface* module. *User Profiler* collects information about the users like their priorities, interest in specific services, etc., so that some special advantages can be given to the users. *Negotiation and Pricing* is subdivided into two parts based on the distinction of their functionality, i.e.,: *Energy-aware Negotiation* and *Service Pricing*. *Energy-aware Negotiation* is responsible for negotiating with the users according to the user's QoS demand and energy-aware technique to settle the service level agreement (SLA) and penalties with specified prices between the cloud provider and user. *Service Pricing*, as the name depicts, deals with pricing issues with users according to the type and scale of service opted. The third module of *User Interface* sub layer, i.e., *Service Request Interface* is subdivided into *Service Request Decoder* and *Service Request Analyzer*. When the service is requested, first it is identified by the *Service Request Decoder* and sent to *Service Request Analyzer* to decide whether the requested service can be granted or not according to the SLA and availability of resource.

Cloud Interface sub layer consists of *Service Request Scheduler*, *Monitors,* and *Energy-aware Virtual Machine Manager* Module. Gathered user's service requests are scheduled according to predefined policy through *Service Request Scheduler* [11]. *Virtual Machine Monitor, Physical Machine Monitor, Performance Monitor, Resource Utilization Monitor,* and *Energy Monitor* are the parts of *Monitors* module which provides its monitored information to the other components of the cloud interface so that they can perform their tasks.

Virtual Machine Monitor and *Physical Machine Monitor* are responsible for probing into system and keeping the status of virtual machines and physical machines, respectively, by counting and identifying which virtual machines and physical machines are on/off. Performance of the user's service request according to the SLA is monitored by the *Performance Monitor*. *Resource Utilization Monitor* interacts with the virtual machines to monitor the amount of resources utilized while processing the service requests. Energy consumption by each physical machine is monitored by *Energy Monitor*.

The third module of *Cloud Interface* sub layer is *Energy-aware Virtual Machine Manager,* which is responsible for saving energy and carbon footprints by allocating new virtual machine request on physical machine consolidating the allocated virtual machine to minimize the current running physical machine and switching off the idle physical machines. It is carried out in two phases by *Virtual Machine Allocation Manager* and *Consolidation Manager*. *Virtual Machine Manager* deals with allocating new request of virtual machine on physical machine. This process of allocation is modeled through classical bin-packing algorithm, e.g., best-fit
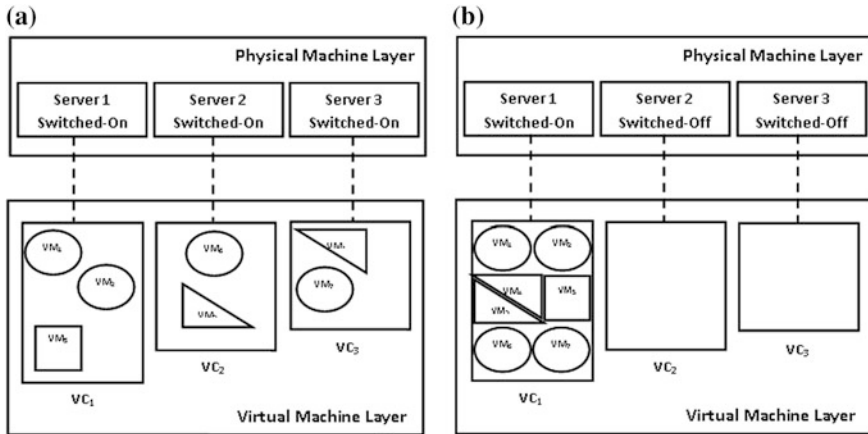
Fig. 3 Consolidation of virtual machines in environment-aware architecture of cloud. **a** Before consolidation. **b** After consolidation

algorithm, first-fit algorithm, best-fit decreasing algorithm, etc. The bin-packing algorithm is analogous to the allocation process, in which packets of given various sizes (to model Virtual Machines) are attempted to be packed into a minimum number of containers/bins (to model physical machines) [12]. The work of *Consolidation Manager* is to consolidate the allocated virtual machine by live migrating them among other available physical machines and switching off the idle physical machine(s) as shown in Fig. 3. Virtualization capacity is the maximum number of virtual machine which can be allocated on an individual physical machine depending on the resource availability of individual physical machine. It is denoted by VC_i and represented by rectangular area for individual physical machines. Virtual machine is denoted by VM_j and is represented by circle, triangle, and square shapes and sizes according to the resource requirement to show heterogeneity.

Consolidation Manager is composed of *Live Migration Manager*, *Consolidation Controller,* and *On/Off-Physical Machine Controller. Live Migration Manager* deals with migrating virtual machine from underloaded or overloaded physical machine. Migration from underloaded physical machine is carried out to save energy while migration from overloaded physical machine is carried out to avoid violation of any QoS. *Live Migration Controller* is responsible for enabling the process of live migration by using various *Live Migration Strategies. Consolidation Controller* is responsible for initiating and terminating consolidation, and its output is used by *On/Off-Physical Machine Controller* to switch off the idle physical machine to save energy and carbon footprints, and to switch-on physical machine for fulfilling new virtual machine requirement.

Physical Machine Layer consists of multiple servers on which various service requests get executed. Virtualization hides the infrastructure complexity of underlying hardware and abstracts the physical infrastructure. It allows creating multiple

virtual machines on single server to improve utilization of server. *Virtual Machine Layers* consist of multiple virtual machines; they are operating system independent and can run multiple applications concurrently on single server to properly utilize the hardware resources.

4 Methodology

The whole method of consolidating virtual machines is carried out in two phases: allocation and optimization. In the first phase, newly requested VMs are placed into suitable physical machines on the basis of CPU utilization using bin-packing algorithms. In this work, two bin-packing algorithms are used: first-fit decreasing and first-fit. In first-fit decreasing algorithm, the requested virtual machines are sorted in nonincreasing order according to their CPU utilization and the suitability of physical machine is checked. The virtual machine is allocated to the first physical machine, which satisfies the criteria. In the first-fit algorithm, similar steps are carried out except sorting of the requested virtual machine. In the second phase, currently allocated VMs are optimized by migrating VMs among physical machines. Migration of VMs depends upon the predefined lower utilization threshold and upper utilization threshold of physical machines. Minimum migration time (MMT) policy is used to select VM for migration [6]. This policy selects that VM which takes minimum time for migration. After selecting the VM for migration, again bin-packing algorithms are used to place them into suitable physical machines.

5 Experiments and Results

The proposed architecture has been evaluated by simulation using CloudSim toolkit. Data center consists of physical machines. Users' requests are submitted through VMs. Two types of physical machines are used consisting of 2660 and 1860 MIPS and both having two processing elements. The lower and upper utilization thresholds are set to 40 and 90 %, respectively. In this experiment, first-fit decreasing and first-fit algorithms with minimum migration time policy are used. For comparative purpose, simulation of non environment-aware policy is also carried out, which does not consolidate VMs, i.e., it does not have the migration of VMs.

Table 1 depicts the simulation result for energy consumption, carbon footprints, SLA violation, and number of migrations. Nonenvironment-aware policy has very high energy consumption in comparison to minimum migration time policy with first-fit decreasing and first-fit algorithm.

The comparison of carbon footprints between non environment-aware policy and minimum migration time policy with first-fit decreasing algorithm is shown in Fig. 4, and between non environment-aware policy and minimum migration time

Table 1 Results

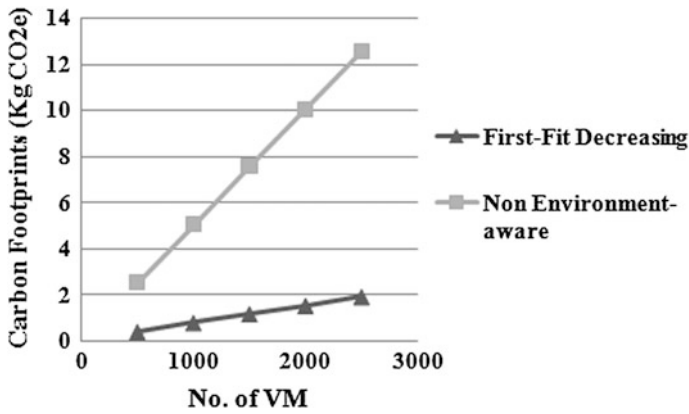| Policy	Energy consumption (kWh)	Carbon footprints (kg CO$_2$e)	SLA violation (%)	No. of migration
Non environment-aware	23.33	12.54		
MMT-first-fit	3.90	2.10	0.32	1375
MMT-first-fit decreasing	3.56	1.91	2.8	1013

Fig. 4 Carbon footprints: first-fit decreasing and non environment-aware
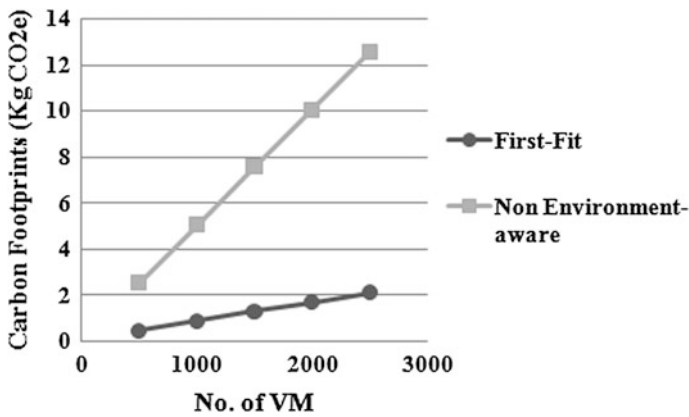

Fig. 5 Carbon footprints: first-fit and non environment-aware

policy with first-fit algorithm is shown in Fig. 5. They show that with migration of VMs carbon emission can be saved, both first-fit decreasing and first-fit algorithms using minimum migration time policy have less carbon footprints in comparison to non environment-aware policy.
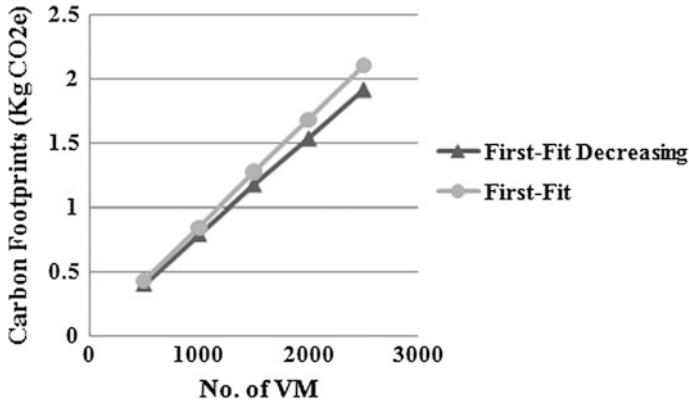
Fig. 6 Carbon footprints: first-fit decreasing and first-fit

On comparing the energy consumption of first-fit decreasing and first-fit algorithm using minimum migration time policy, it is evaluated that first-fit decreasing algorithm consumes less energy than first-fit algorithm, as depicted in Fig. 6.

From Table 1, it is observed that by using minimum migration time policy with first-fit decreasing algorithm 84 % carbon footprints can be saved with respect to non environment-aware policy and giving 2.8 % SLA violation. While minimum migration time policy with first-fit algorithm saves 83 % carbon footprints with respect to non environment-aware policy and giving 0.32 % SLA violation.

6 Conclusion

Carbon footprints are estimated for the proposed architecture obtained by extending the existing classical cloud architecture. This novel architecture is obtained by adding an energy-aware service provider layer. Live migration applied in this proposed architecture minimizes the number of running physical machines to minimize the required energy. The standard bin-packing algorithms; first-fit decreasing and first-fit used along with live migration save 84 and 83 % carbon footprints, respectively. As a future work, impact on other factors like propagation delay, server's temperature, etc., can be studied. Virtual machine allocation can be done using other resources like RAM, disk, bandwidth, etc., instead of using only CPU utilization.

References

1. Mell, P., Grance, T.: The NIST Definition of Cloud Computing. Technical Report. National Institute of Standards and Technology, USA (2011).
2. Delforge, P., Whitney, J.: Data Center Efficiency Assessment. Technical Report. Natural Resource Defence Council, New York (2014).
3. Calheiros, R. N., Ranjan, R., Beloglazov, A., Rose, C. A. F. D., Buyya, R.: CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. Software- Practice & Experience. vol. 21, issue. 1. John Wiley & Sons Ltd, New York (2011).
4. Department for Environment Food & Rural Affairs, http://www.ukconversionfactorsc arbonsmart.co.uk/.
5. Verma, A., Ahuja, P., Neogi, A.: pMapper: Power and Migration Cost Aware Application Placement in Virtualized Systems. In: Issarny, V., Schantz, R. (eds.) Middleware 2008. LNSC, vol. 5346, pp. 234–264. Springer, Heidelberg (2008).
6. Buyya, R., Beloglazov, A.: Optimal Online Deterministic Algorithms and Adaptive Heuristics for Energy and Performance Efficient Dynamic Consolidation of Virtual Machines in Cloud Data Centers. Concurrency and Computation: Practice and Experience. vol. 24, issue. 13, pp: 1397–1420. John Wiley & Sons Ltd, UK (2012).
7. Rajabi, A., Ebrahimirad, V., Yazdani, N.: Decision Support-as-a-Service: An Energy-aware Decision Support Service in Cloud Computing. In: 5th Conference on Information and Knowledge Technology, pp. 71–76. IEEE Press, Iran (2013).
8. Li, Y., Wang, Y., Yin, B., Guan, L.: An Energy Efficient Resource Management Method in Virtualized Cloud Environment. In: 14th Asia-Pacific Network operations and management symposium, pp. 1–8. IEEE Press, South Korea (2012).
9. Wang, J., Huang, C., He, K., Wang, X., Chen, X., Qin, K.: An Energy-aware Resource Allocation Heuristics for VM Scheduling in Cloud. In: IEEE International Conference on High Performance Computing and Communications & IEEE 10th International Conference on Embedded and Ubiquitous Computing, pp. 587–594. IEEE Press, China (2013).
10. Clark, C., Fraser, K., Hand, S., Hansen, J.G., Jul, E., Limpach, C., Pratt, I., Warfield, A.: Live Migration of Virtual Machines. In: 2nd USENIX Symposium on Network Systems Design and Implementations, pp. 273–286. USENIX Association, Berkeley (2005).
11. Brucker, P.: Scheduling Algorithm, 5th ed., Springer-Verlag Berlin Heidelberg (2007).
12. Coffman, E.G., Csirik, J., Woeginger, G.J.: Approximate Solutions to Bin Packing Problems. In: Applied Optimization, Pardalos, P.M., Resende, M.G.C (eds.), Handbook of Applied Optimization, Oxford University Press, pp. 607–615, New York (2002).