

A Review of Energy-aware Cloud Computing Surveys

Azlan Ismail*, Nor Adzlan Jamaludin, Suzana Zambri

Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA (UiTM), 40450, Shah Alam, Selangor, Malaysia
telp/fax: 603-55211183/603-55435501

*Corresponding author, e-mail: azlanismail@tmsk.uitm.edu.my

Abstract

The increasing demands on the usage of data centers especially in provisioning cloud applications (i.e. data-intensive applications) have drastically increased the energy consumption and becoming a critical issue. Failing to handle the increasing in energy consumption leads to the negative impact on the environment, and also negatively affecting the cloud providers' profits due to increasing costs. Various surveys have been carried out to address and classify energy-aware approaches and solutions. As an active research area with increasing number of proposals, more surveys are needed to support researchers in the research area. Thus, in this paper, we intend to provide the current state of existing related surveys that serve as a guideline for the researchers as well as the potential reviewers to embark into a new concern and dimension to compliment existing related surveys. Our review highlights four main topics and concludes to some recommendations for the future survey.

Keywords: classification, cloud computing, energy-aware, surveys

Copyright © 2018 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction

The emergence and development of complex data-intensive applications influence the need to increase and improve the service provisioning of cloud infrastructure. A few studies have shown that the clouds especially its data center consumed lots of energy. Many internet companies (e.g. Amazon, Google) are operating such huge data centers around the world [1]. This huge demand on the data center causes high energy consumption from different cloud elements that include software, hardware (i.e. IT equipment and non-IT equipment), and network. The effect of high energy consumption can be interpreted as high operational cost, which reduces the profit margin of cloud providers, and leads to high carbon emissions which are not environmentally friendly [1]. As a consequence, the environment is in danger due to the high amount of carbon emissions. Thus, there is a need to overcome this issue with energy-efficient solutions which can minimize the impact of cloud computing on the environment.

Cloud computing needs energy efficiency solutions in managing resources (i.e. Virtual machines (VMs), physical servers). From the literature, various approaches and techniques have been proposed to realize energy efficiency of cloud computing. From the opposite perspective, the cloud computing itself can be viewed as the most desirable technology to deal with the energy crisis [2]. This fact is supported by several studies, among others [3] that predict the energy consumption may decrease by 87% if organizations move to the cloud.

Many surveys have been done in relation to energy-aware cloud computing. Thus, there is a need to understand the current state of these surveys to assist the future survey direction. This is significance to minimize from repeating similar surveys and to enrich the information gathering and analysis of energy-aware cloud computing. Therefore, this paper aims to contribute to the classification of existing surveys and several recommendations for the direction of future survey.

The rest of this paper is organized as follows. We introduce the key concepts related to this paper in section 2. We discuss and summarize existing energy-aware cloud surveys with a working taxonomy in section 3. We provide several suggestions for future survey in section 4. We conclude the paper in section 5.

2. Background

In this section, we introduce two main concepts related to this paper, namely, cloud computing and energy-aware cloud management.

2.1. Cloud Computing

Cloud computing can be defined as a paradigm for the dynamic provisioning of computing services provided by data centers that commonly utilize Virtual Machine (VM) technologies [4]. The definition given by U.S. National Institute of Standards and Technology (NIST) states that the cloud computing is a model that enables ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources, e.g., networks, servers, storage, applications, and services [5].

The computing services provided by cloud computing to the cloud consumers include Infrastructure as a service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) [6]. The IaaS layer is responsible for managing the physical machines, creating a pool of virtual computation or storage resource. The PaaS layer consists of operation systems and application frameworks. Meanwhile, the SaaS layer contains the actual cloud application. Furthermore, the IaaS can be divided into three sub-layers. Firstly, the physical resource layer that consists of a traditional data center which contains IT equipment (e.g. servers, network devices, disks) and non-IT equipment (e.g. cooling, lightening, air conditioner). Secondly, the virtual resource layer that consists of multiple virtualized computation or storage resources provided through virtualization technology. Thirdly, the management tool layer that provides the capability to manage virtual resources, accounting, and monitoring. In addition, some researchers have further classified the cloud services into Hardware-as-a-Service (HaaS) and Data-as-a-Service(DaaS) [7].

There are a set of key characteristics of clouds that includes virtualization, elasticity, service-orientation, dynamic and distributed, shared, market-oriented and autonomic [1]. Virtualization is a concept of provisioning resources to enable cloud to deploy and execute many applications on possibly different operating systems and running on the same physical machines. The resource refers to the server, storage and network [8]. The elasticity provides the support of scaling capability, which is to scale up or down the cloud resources depending on the workload demands. Implementing Service-oriented based on Service-oriented Architecture (SOA) enables various services can be offered from different layers. Dynamic and distributed emphasizes the ability of clouds to delivery services in high performance. The shared characteristics highlight the shared infrastructure of clouds based on the multi-tenant model. The market-oriented represents the common practice of utilizing cloud services (also called metered services) as a pay-per-use basis. The autonomic represents the need for highly reliable cloud services which can manage itself and recover from failures.

2.2. Energy-aware Cloud Management

Efficient energy management is one of the most challenging research areas in resolving the issue of increasing power consumption. Various studies have focused on the data centers since it requires large amounts electrical energy that contributes to high operational costs and carbon footprints to the environment [4]. Any efforts and solutions which can significantly reduce energy consumption for the data centers and would also make a significant contribution to greater environmental sustainability [6]. However, striving to minimize the energy consumption is insufficient as it can potentially affect other objectives, such as QoS and scalability. Thus, another major challenge is to tackle the tradeoffs between multi-objectives. Despite these challenges, it has been known that the virtualization technology provides great opportunities to improve energy efficiency for the cloud data center as well as satisfying other objectives. Many strategies have been proposed in relation to effectively and efficiently handle virtualization [6]. Among the strategies are energy-aware dynamic resource allocation, thermal-aware energy efficient load balancing, and energy-aware task consolidation [9] as well as VM consolidation [10].

3. Review of Existing Surveys

In this section, we review the existing energy-aware cloud surveys according to four topics. Through this review, we propose a classification of existing surveys as illustrated in

Figure 1. For each topic, we provide the highlighted dimensions from the reviewed papers. The topics are discussed as follows.

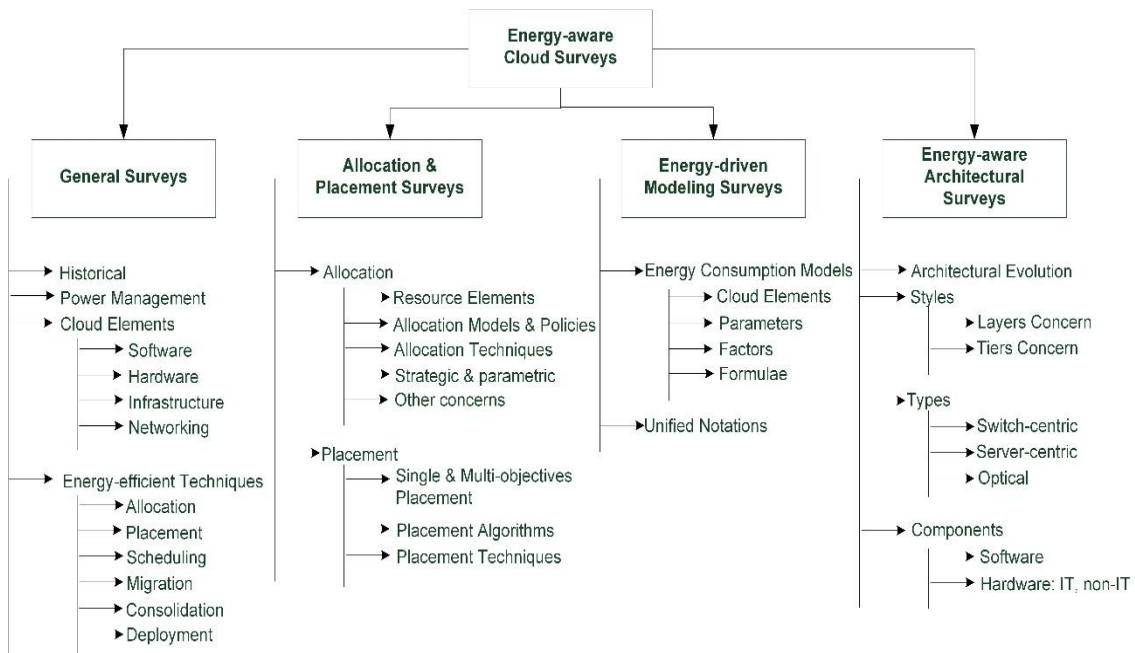


Figure 1. Classification of energy-aware cloud computing surveys

3.1. General Energy-aware Cloud Surveys

We have selected a set of surveys under this topic, namely, [1] [6] [11–16]. We now elaborate each of them as follows. Surveying the historical perspective can assist to understand the motivation and evolution of energy-aware solutions. For this reason, the interested reader can refer to Kaur and Chana [14]. The historical evolution is discussed from data centers perspective. It has shown that ICT industry becomes the major culprit of energy consumption throughout the world. ICT industry also contributes to increasing Greenhouse Gases which is posing a danger to environmental sustainability. The maximal power consumption of ICT is also caused by a large number of servers hosted in data centers where some of them remain idle most of the time and consume energy.

The next aspect is the power management which becomes the key aspect for an energy-aware solution. Two of the selected papers have surveyed into this aspect. Firstly, Beloglazov et al. [11] that presented a set of taxonomies for power management of cloud computing. The higher level of computing systems taxonomy focuses on the Static Power and Dynamic Power Management. The taxonomy for the hardware level is divided into Dynamic Component Deactivation and Dynamic Performance Scaling. The taxonomy of the operating system level is classified according to several dimensions (i.e. application adaptation, target systems and power saving techniques) as well as the data center level (i.e. virtualization, system resources and workload).

Secondly, Raj and Shriram [15] that emphasized on the joint IT and non-IT components. A classification of power management approaches has been presented from several aspects, among others, the application, cluster architecture, and server hardware and other infrastructure level.

Other surveys namely [1], [6], [12,13], [16] have presented different interpretation of cloud elements. The survey by Kumar and Buyya [1] emphasized the key characteristics exhibited by clouds, the components of clouds, the deployment models through public, private and hybrid clouds, and energy usage model. Furthermore, they explained the key features of clouds to enable Green computing, namely, dynamic provisioning, multi-tenancy, server utilization, and data center efficiency. Then, the state-of-the-art approaches towards energy

efficiency are discussed followed with a proposal of green cloud architecture. The key ingredient of the architecture is the application of green aware for each layer of cloud, SaaS, PaaS and IaaS.

Another survey by Jing et al. [6] focused on the cloud elements categorized as IT (e.g. processor, server, storage, network) and non-IT equipment (cooling, lightning, air conditioning). These elements were then used to classify the existing techniques. Among the techniques discussed for the IT-equipment are the task scheduling [17], the voltage scaling, and the VM migration/allocation/provisioning. Meanwhile, the techniques related to the non-IT equipment involves efficient thermal management which can be divided into two levels, facility and system level. The facility level concerns with optimizing the flow of hot and cold air in the data center. The system level addresses the optimization of placing the workload across the servers to minimize the cooling requirement.

Mastelic et al. [12] viewed the cloud elements from the infrastructure domains with two categories, hardware and software. The hardware part refers to IT-related equipment i.e. network and servers. The energy consumption within the network context refers to the connections inside of a data center, the fixed network between data centers, and the end user network. The server domain includes computing and storage servers, as well as other components such as processors, memory, etc. Meanwhile, the software part consists of Cloud Management Systems (CMS) and appliances.

Meanwhile, Moghaddam et al. [13] addressed the networking elements which have a huge influence to the cloud environment, specifically, the network devices which consume around 90% even at the low load or in an idle state [18]. A metamodel has been proposed to classify the existing works that include strategy, scale, solution, technology, and energy efficiency evaluation. The strategy refers to the existing proposed plan for total energy consumption reduction. The scale refers to the size of the data center, namely the intra-data center, inter-data center, and mixed-scale data center. The solution focuses on the device, routing/switching protocols, network architecture, and decision framework. The technology summarizes the technologies to realize the solutions. Meanwhile, the energy efficiency evaluation focuses on the proposed evaluation methods which can be either through simulations, empirical experiments, and numerical analysis.

Finally, Zakarya and Gillam [16] paid a special attention to the ICT equipment in large scale computing systems (i.e. High Performance Computing (HPC), Cluster, Grid, Cloud, and Datacenter). The studies also include the modeling aspect of energy consumption for non-virtualized and virtualized platforms, and workload scheduling. They presented a taxonomy of energy efficient computing from hardware, resource and application perspective. They then discussed the existing energy efficient techniques based on host, network, thermal-aware, virtualized and geographic level.

3.2. Energy Efficient Resource Allocation and Placement Surveys

In this section, we have chosen [19] and [20] for the energy-aware resource allocation, and [21] and [22] for the VM placement techniques with energy efficiency concern. We now briefly summarized each of the work. In Hameed et al. [19], they reviewed the existing energy efficient resource allocation from hardware-based and software-based techniques. They classified the existing proposals from four dimensions and presented the advantages and disadvantages of each techniques. The dimensions are resource adaptation policy, objective function, allocation method, allocation operation, and interoperability.

Meanwhile, Madni et al. [20] conducted a systematic review by focusing on the cloud infrastructure layer that includes energy-aware concern. The review include the categorization and discussion of existing works and discussion of the addressed resource elements and parameters in the existing studies. The categorization is based on two core dimensions, strategic (i.e. artificial intelligent, dynamic, predictive) and parametric (i.e. cost, performance, load balance, energy-aware, QoS, utilization) resource allocation.

In the case of placement and migration, the work by Filho et al. [21] considered the proposed techniques in achieving multi-objectives and conflicting goals that includes energy saving goal. They discussed the state-of-the-art of placement and migration techniques based on a few categories, among others, business-aware, energy-aware, interference-aware, and fault-tolerance and elasticity-aware. Then, four types of reviews are presented. Firstly, the architecture, openness of source code, pro-activeness, and problem formulation and

technologies. Secondly, the concerns of each proposed works. Thirdly, the experiments and implementation approach. Fourthly, the paper category and involved parameters.

Attaoui and Sabir [22] also reviewed the VM placement techniques, but focussed on the multi-objective optimization that includes energy saving. They discussed the state-of-the-art works from five objective dimensions, namely, energy, cost, network traffic, resource and QoS. They also presented the existing proposals from the algorithm types, notably, heuristic, meta-heuristic, deterministic, and approximation algorithms. The findings of the review stated that heuristic and/or metaheuristics are dominant selection since it can provide good quality solutions.

3.3. Energy-aware Modeling Surveys

In this section, we have chosen the recent works, namely [23] and [24] which specifically focus on the energy-related modeling. The motivation of the modeling approach is to represent real-world systems and thus can assist to quantify some values. Each of them is summarized as follows.

Dayarathna et al. [23] surveyed the state-of-the-art techniques that targeted at the infrastructure layer. They presented the energy consumption modeling and prediction process, and the system perspective of data center energy consumption. Then, a set of modeling approaches is discussed, namely, digital circuit power, aggregate view of server power, processor power, memory and storage power, data centers power, and software energy. All of these approaches are summarized and classified according to software-centric and hardware-centric classification.

Li Z [24] came up with a survey of the existing modeling efforts by focusing on the energy consumption of cloud applications. A set of factors were identified related to the environment and workload that can influence the energy consumption on cloud applications. Then, they presented a model with unified notations which can assist to understand the fundamental concepts and facilitate simulations to address cloud application energy efficiency problems.

3.4. Energy-aware Architectural Surveys

In this section, we have chosen two surveys, namely [25], and [26] that reviewed on the energy-aware architectures. Each of them is summarized as follows. Hammadi and Mhamdi [25] surveyed the architectural evolution of data center networks and their energy efficiency. They provided a classification of existing data center network architectures into two broad classifications, namely, electronic and optical data center. The electronic data center is further divided into two topologies. Firstly, the conventional data center called the switch-centric. Secondly, the emerging architecture that is based on packet-switched electronic networks, called the server-centric. Meanwhile, the optical data center mainly relies on a mixture of active and passive optical devices to provide switching, routing, and interconnection.

Ruiu et al. [26] also categorized the data center architectures into switch-centric or server-centric. For the analysis, they considered two and three-tier architectures. Furthermore, a comparison study is reported that includes different power models and the resource allocation scheme. The power models focused on energy-proportional, non-proportional and realistic. Meanwhile, the focused allocation schemes are Random Server Selection (RSS) and Min-Network Power (MNP). A testing framework is also illustrated to investigate the performance of different data center architectures with power consumption profiles.

4. Recommendations

The need for different aspects of future surveys are important to foster the energy-aware cloud research. Therefore, we suggest four potential areas. Firstly, the aspect of survey on *energy-aware multi-controllers*. The current surveys are lacking in terms of reviewing the multi-controllers perspective of energy-aware clouds. This kind of survey is relevant since there are a few approaches aligned with this direction. For instance, the work by Hasan et al. [27] that proposed an auto-scaler architecture based on Monitor-Analyze-Plan-Execute-Knowledge (MAPE-K) loop framework [28].

Secondly, the area of survey related to *the renewable energy*. The current surveys are lacking in terms of proposals that take renewable energy into consideration. For instance,

Hasan et al. [29] that presented a concept of virtualization of green energy. Specifically, they introduced a scheme for green energy management in the presence of renewable energy in the data center. Therefore, a distinct classification is needed for the related works.

Thirdly, the survey related to *energy-aware resiliency*. We notice a branch of study in this direction, known as *cloud resiliency* that refers to the ability of the cloud systems to remain reliable, survivable, and dependable in case of failures. These failures may originate from any of the major components in a cloud architecture which can affect others. We note a survey done by Colman-Meixner et al. [30]. To compliment the existing surveys, the future survey can focus on investigating cloud resiliency approaches with energy-aware concern.

Finally, the aspect of survey on *energy-aware prediction*. Prediction plays an important role to enable proactive solution for energy-aware cloud computing. A survey by Mukwehwo and Celik [31] addressed the fault-tolerance of cloud systems that includes some prediction works. However, this survey does not take the energy-aware into consideration. An example of predictive solution that relates to the energy-aware is referred to Farahnakian et al. [10]. This survey should consider prediction from multi-criteria aspects, such as energy, performance, workload, and cost.

5. Conclusion

Energy-aware cloud computing is a significant research area towards realizing sustainable and green cloud computing. For this reason, this paper contributes to the analysis of existing related surveys and proposals of energy efficiency techniques. Based on our analysis, we have identified four main topics to classify the existing surveys namely, (i) general energy-aware cloud surveys, (ii) energy-efficient resource allocation and placement surveys, (iii) energy-driven modeling surveys, and (iv) energy-aware architectural surveys.

We have also recommended several directions which can be taken into future energy-aware cloud reviews. These recommendations were formulated by considering the established survey classification and the recent research works. Firstly, we recommend a survey on the aspect of energy-aware cloud multi-controllers. Secondly, we suggest a review on the virtualization of green energy. Thirdly, we recommend a survey on energy-aware cloud resiliency.

Acknowledgement

This research has been supported by a Research Grant, 600-IRMI/ PERDANA 5/3 BESTARI (048/2018), funded by Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia.

References

- [1] S Kumar and R Buyya. Green Cloud Computing and Environmental Sustainability. in *Harnessing Green It*, Chichester, UK: John Wiley & Sons, Ltd. 2012: 315–339.
- [2] F Owusu and C Pattinson. *The Current State of Understanding of the Energy Efficiency of Cloud Computing*. in 2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications. 2012: 1948–1953.
- [3] Cloud computing saves energy on huge scale, says new study | Envantage. Envantage. [Online]. Available: <http://www.envantage.co.uk/cloud-computing-saves-energy-on-huge-scale-says-new-study.html>. Accessed May 12, 2018.
- [4] A Beloglazov, J Abawajy, and R Buyya. Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing. *Futur. Gener. Comput. Syst.* 2012; 28(5) 755–768.
- [5] P Mell and T Grance. The NIST Definition of Cloud Computing. *NIST Pubs.* 2009. Special Publication (NIST SP).
- [6] SY Jing, S Ali, K She, and Y Zhong. State-of-the-art research study for green cloud computing. *J. Supercomput.* 2013; 65(1): 445–468.
- [7] L Wang et al. Cloud Computing: a Perspective Study. *New Gener. Comput.* 2010; 28(2): 137–146.
- [8] O Sri Nagesh, T Kumar, V Venkateswararao, and OS Nagesh. A Survey on Security Aspects of Server Virtualization in Cloud Computing. *Int. J. Electr. Comput. Eng.* 2017; 7(3): 1326–1336.
- [9] MK Gourisaria, SS Patra, and PM Khilar. Minimizing Energy Consumption by Task Consolidation in Cloud Centers with Optimized Resource Utilization. *Int. J. Electr. Comput. Eng.* 2016; 6(6): 3283–3292.

- [10] F Farahnakian, T Pahikkala, P Liljeberg, J Plosila, NT Hieu, and H Tenhunen. Energy-aware VM Consolidation in Cloud Data Centers Using Utilization Prediction Model. *IEEE Trans. Cloud Comput.* 2016; Early Access: 1–1.
- [11] A Beloglazov, R Buyya, YC Lee, and A Zomaya. A Taxonomy and Survey of Energy-Efficient Data Centers and Cloud Computing Systems. *Adv. Comput.* 2011; 82: 47–111.
- [12] T Mastelic, A Oleksiak, H Claussen, I Brandic, JM. Pierson, and AV Vasilakos. Cloud Computing: survey on energy efficiency. *ACM Comput. Surv.* 2014; 47(2): 1–36.
- [13] FA Moghaddam, P Lago, and P Grosso. Energy-Efficient Networking Solutions in Cloud-Based Environments. *ACM Comput. Surv.* 2015; 47(4): 1–32.
- [14] T Kaur and I Chana. Energy Efficiency Techniques in Cloud Computing. *ACM Comput. Surv.* 2015; 48(2): 1–46.
- [15] VKM Raj and R Shriram. Power management in virtualized datacenter – A survey. *J. Netw. Comput. Appl.* 2016; 69: 117–133.
- [16] M Zakarya and L Gillam. Energy efficient computing, clusters, grids and clouds: A taxonomy and survey. *Sustain. Comput. Informatics Syst.* 2017; 14: 13–33.
- [17] KK Chakravarthi and V Vijayakumar. Workflow Scheduling Techniques and Algorithms in IaaS Cloud: A Survey. *Int. J. Electr. Comput. Eng.* 2018; 8(2): 853.
- [18] B Heller, S Seetharaman, and P Mahadevan. *ElasticTree: Saving Energy in Data Center Networks*. Proceeding NSDI'10 Proceedings of the 7th USENIX conference on Networked systems design and implementation. 2010: 17-17.
- [19] A Hameed *et al.* A survey and taxonomy on energy efficient resource allocation techniques for cloud computing systems. *Computing.* 2016; 98(7): 751–774.
- [20] SHH Madni, MSA Latiff, Y Coulibaly, and SM Abdulhamid. Recent advancements in resource allocation techniques for cloud computing environment: a systematic review. *Cluster Comput.* 2017; 20(3): 2489–2533.
- [21] MC Silva Filho, CC Monteiro, PRM Inácio, and MM Freire. Approaches for optimizing virtual machine placement and migration in cloud environments: A survey. *J. Parallel Distrib. Comput.* 2018; 111: 222–250.
- [22] W Attaoui and E Sabir. Multi-Criteria Virtual Machine Placement in Cloud Computing Environments: A literature Review. 2018.
- [23] M Dayarathna, Y. Wen, and R Fan. Data Center Energy Consumption Modeling: A Survey. *IEEE Commun. Surv. Tutorials.* 2016; 18(1): 732–794.
- [24] Z Li *et al.* A Survey on Modeling Energy Consumption of Cloud Applications: Deconstruction. State of the Art, and Trade-Off Debates. *IEEE Trans. Sustain. Comput.* 2017; 2(3): 255–274.
- [25] A Hammadi and L Mhamdi. A survey on architectures and energy efficiency in Data Center Networks. *Comput. Commun.* 2014; 40: 1–21.
- [26] P Ruiu, A. Bianco, C Fiandrino, P Giaccone, and D Kliazovich. *Power comparison of cloud data center architectures. in 2016 IEEE International Conference on Communications (ICC).* 2016: 1–6.
- [27] MS Hasan, F Alvares, T Ledoux, and JL. Pazat. Investigating Energy Consumption and Performance Trade-Off for Interactive Cloud Application. *IEEE Trans. Sustain. Comput.* 2017; 2(2): 113–126.
- [28] JO Kephart and DM Chess. The vision of autonomic computing. *Computer (Long. Beach. Calif).* 2003; 36(1): 41–50.
- [29] MS Hasan, Y Kouki, T Ledoux, and JL Pazat. Exploiting Renewable Sources: When Green SLA Becomes a Possible Reality in Cloud Computing. *IEEE Trans. Cloud Comput.* 2017; 5(2): 249–262.
- [30] C Colman-Meixner, C Develder, M Tornatore, and B. Mukherjee. A Survey on Resiliency Techniques in Cloud Computing Infrastructures and Applications. *IEEE Commun. Surv. Tutorials.* 2016. 18(3): 2244–2281.
- [31] MA Mukwevho and T Celik. Toward a Smart Cloud: A Review of Fault-tolerance Methods in Cloud Systems. *IEEE Trans. Serv. Comput.* 2018: 1–1.