Volume 5, Issue 10, October-2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Survey on Energy Efficiency in Cloud Computing

Ranjana Badre
Department of Computer Engineering
MIT Academy of Engineering, Alandi, Pune
University of Pune, Maharashtra, India

Abstract: ***Todays most emphasized Information and Communications Technology (ICT) paradigm is Cloud computing. This is because cloud computing supports a great infrastructure that includes large data centers comprising thousands of server units and other supporting equipment. Their share in power consumption generates between 1.1% and 1.5% of the total electricity used worldwide and is projected to rise even more. So in this paper a comprehensive analysis of an infrastructure supporting the cloud computing paradigm with regards to energy efficiency is presented. A systematic approach for analyzing the energy efficiency of most important data center domains, including server and network equipment, as well as cloud management systems and appliances consisting of a software utilized by end users is presented. From this existing challenges are extracted are highlighted.***

Keywords: Cloud computing, energy efficiency, data center, network. Equipment, Cloud management system

I. INTRODUCTION

Today the technological machinery is easily accessible to the average person. But due to its global usage, technological machinery creates an increasing demand for more energy. From 1990 until today, power consumption doubled from 10k TWh up to 20k TWh worldwide [Enerdata 2014]. Future projections estimate almost 40k TWh by 2040—a 2.2% increase per year [EIA 2013].

To enhance sustainability of the energy supply and to reduce emissions of greenhouse gases and other pollutants, the European Commission pointed out energy efficiency as the most cost effective way for achieving long-term energy and climate goals. ICT has already been recognized as an important instrument for achieving these goals [EU 2008]. However, ICT is also recognized as one of the major energy consumers through equipment manufacture, use, and disposal [Advisory Group 2008], which also became one of the key issues of the Digital Agenda for Europe issued by the European Commission in 2010 [EU 2010].

In this survey, the focus is on energy efficiency of the ICT equipment separated into two domains: Server and Network. The software solutions running on top of ICT equipment; these include the Cloud Management System (CMS) domain for managing a cloud infrastructure and the Appliance domain that represents a software for servicing users are also covered.

For the purpose of the survey, taxonomy and terminology used throughout the article describing energy efficiency in general is defined. It is also applied to the cloud computing infrastructure to create a systematic approach for analyzing the energy efficiency of ICT equipment within a data center.

II. TAXONOMY AND TERMINOLOGY

A. Cloud Computing

Cloud computing is a novel and promising paradigm for managing and providing ICT resources to remote users.It utilizes technologies such as virtualization, distributed computing, Service Oriented Architecture (SOA), and Service Level Agreements (SLAs) [Foster et al. 2008], based on which different service types are offered. As defined by NIST [Mell and Grance 2009], cloud computing recognizes three service models: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS).These services are powered by large data centers comprised of numerous virtualized server instances and high-bandwidth networks, as well as of supporting systems such as cooling and power supplies. The listed equipment can be classified into twotypes, as shown in Figure 1; namely, hardware and software equipment [HoelzleandBarroso 2013].

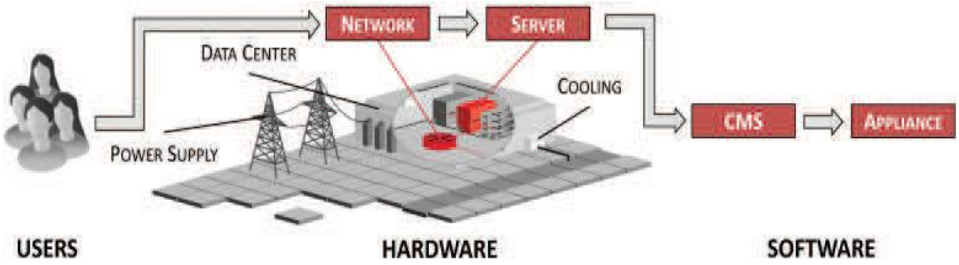

Figure 1 Cloud computing data center domains

Hardware includes both ICT equipment and supporting equipment within a data center. ICT equipment includes *Network* and *Server* domains because they perform the main task of the data center. While supporting equipment's are Power supply, Cooling, and the Data center building itself

Software equipment within a data center includes everything that runs on top of the ICT equipment. It includes two domains *CloudManagement Systems (CMS)* that are used to manage the entire data center and *Appliances*, which include software used by a user.

B. Energy Efficiency

Energy efficiency refers to a reduction of energy used for a given service or level of activity, as defined by theWorld Energy Council [Moisan and Bosseboeuf 2010].However, defining the energy efficiency of data center equipment is extremely difficult [Fanara2007] because it represents a complex system with a large number of components from various research areas such as computing, networking, management, and the like.

Beloglazov et al. [2011] defined an energy model through static and dynamic power consumption, which deals only with energy waste while running idle. On the other hand, Avelar et al. [2012] defined a difference between energy used by ICT and auxiliary equipment in order to measure energy losses.

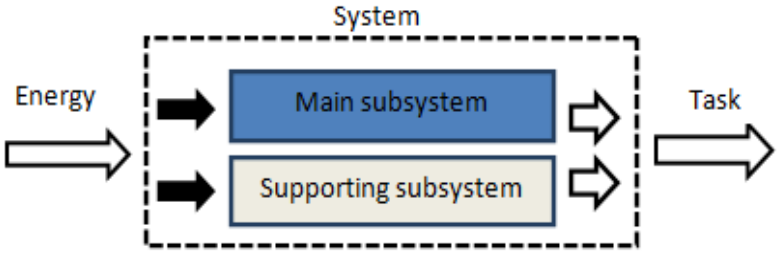

Figure 2 A system and (sub) systems

Figure 2 shows an arbitrary system as a set of interconnected components, in which each component can be observed as a different (sub)system. Therefore, every (sub)system can be optimized for itself, which can affect the energy efficiency of other related systems.Each system requires an input energy for performing acertain task, where a task is an abstract assignment that the system has to perform to fulfill its purpose. To improve the energy efficiency of a system, first it is necessary to identify problems that degrade efficiency.

Figure 3 shows critical points within a system where energy is lost or wasted. Energy loss refers to energy brought to the system but not consumed for its main task (e.g., energy lost due to transport and conversion). This also includes energy used by supporting subsystems, such as cooling or lighting within a data center whose main task is the provision of cloud services. *Energy waste* refers to energy used by the system's main task but without useful output (e.g., energy used while running in idle mode).

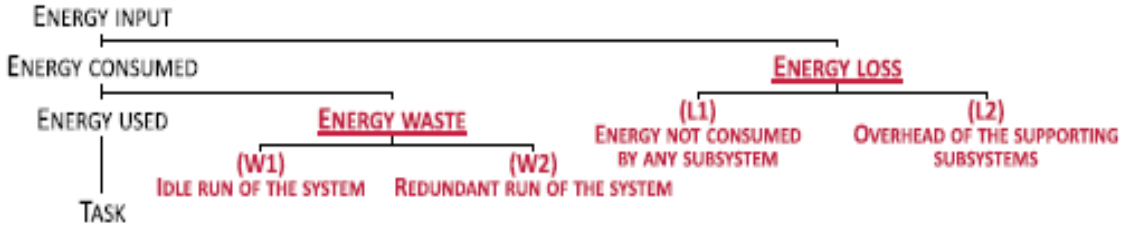

Figure 3 Critical points within a system where energy is lost or wasted

The energy efficiency can be improved by reducing the energy loss and energy waste as follows:

- L1 By implementing more efficient components (e.g., using more efficient power supply units for servers that leak less energy), a percentage of input **energy that is not consumed by a subsystem can be minimized**
- L2 Overhead of supporting system can be reduced by implementing a single cooling unit for the entire cabinet instead of cooling each rack server separately.
- W1 By reducing idle run of the system and increasing utilization or achieve zero energy consumption when no output is produced.
- W2 Energy consumption can be minimized where system performs redundant operations by implementing smart functions and subsystems such as implementing an optimized algorithm that does not require redundant steps to perform the same task.

C. Network

For cloud computing network is key enabling component because it allows communication between computing and storage resources and allows the end user to access them. The energy consumption of the Network domain consists of three main systems: the connections inside of a data center, the fixed network between data centers, and the end user network that increasingly provides the wireless last hop to end users who access services via smart phones, tablets, and laptops. The figure 4 shows energy wastes and losses in the system.

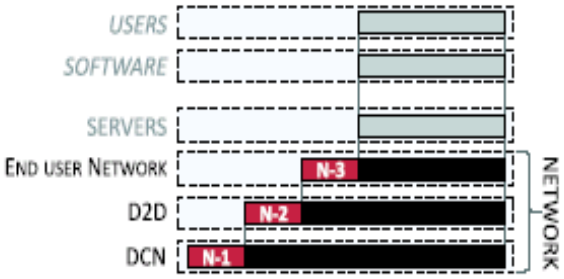

Figure 4 Losses and wastes of Network domain

To reduce energy loss and waste, a number of actions can be taken

- **L1**Reducing the heat load of network equipment inside a data center (N-1). This would reduce its energy consumption and the consumption of its cooling subsystem as well. This can be achieved by adapting the network equipment design suggested by ASHRAE [2012] implementing front to rear air flow.
- **L2**by reducing heat load, a smaller cooling subsystem can be installed, which consumes less energy. Although it comprises basic network equipment, failure handling supported by redundant equipment can also be considered a subsystem because it does not perform the main task of the system. Therefore, moving away from a traditional 2N tree topology toward the more flexible topologies currently being adopted by new data centers, such as Fat-Tree [Al-Fareset al. 2008], BCube [Guo et al. 2009], and DCell [Guo et al. 2008], can provide benefits in terms of improved energy-efficient traffic management.
- **W1**Today's network equipment is not energy proportional, and simply turning on switch can consume over 80% of its max power [Mahadevan et al. 2009]. By implementing power saving modes [Gupta and Singh 2007a; Claussen et al. 2010;Razavi and Claussen 2012], rate adaptation [Lopez-Perez et al. 2014; Gunaratneet al. 2008], or simply turning off unused ports, links, and switches inside a datacenter (N-1) [Heller et al. 2010] would reduce idle energy consumption.

D. Servers

The Server domain includes computing and storage servers [Warkozek et al. 2012], as well as other components such as processors, memory, cabinets. It also considers aspects such as component layout within a rack and component architecture. In a perfect data center, the Server domain, along with the Network domain, would consist only of hardware equipment that consumes energy. Therefore, an obvious goal of every data center owner is to reduce the consumption of all supporting hardware equipment because it represents an energy loss. However, energy loss and waste do not stop there since servers can also contribute to energy waste due to poor server equipment usage policy, as well as energy loss due to a poor energy supply and internal subsystems. As shown in Figure 5, systems of the Serverdomain include the server enclosure, such as server cabinets. Server racks represent another system, and components within a rack, such as CPU, memory, hard-disk, and the like, are the third system.

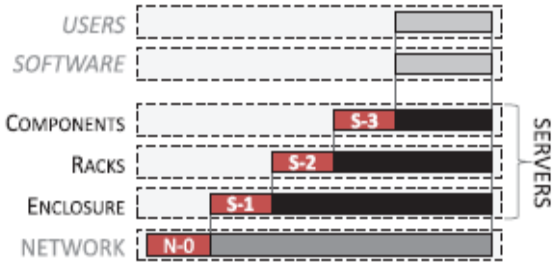

Figure 5 Losses and wastes of server domain

Enclosure (S-1): Enclosures may differ depending on the type of cooling applied to a data center. The most common air-based cooling, based on Computer Room Air Conditioners (CRACs), requires enclosures to have air inlets and outlets on opposite sides. The second type of cooling is indirect liquid cooling. Chilled water is delivered to the enclosure where it is used to absorb heat from the air that is used to cool servers. The enclosure can contain a closed loop of air or implement rear-door (or side-door) cooling, in which the cooled air is pushed back into the server room. Finally, direct liquid cooling solutions have been recently gaining interest [Haywood et al. 2012].This type of cooling is particularly efficient for powerful and heavily loaded servers, as in High Performance Computing (HPC) applications; however, it may be also useful for cloud infrastructures. In enclosures with direct liquid cooling, warm water is used to cool server components directly, most commonly through the use of cold plates [Coolit 2013] or microchannels [IBM 2013]. Recently, other approaches based on the immersion of the whole server in a dielectric fluid have emerged (e.g., Iceotope system [Icetope 2013]). Liquid cooling approaches provide significant energy savings (up to around 40% compared to air-based cooling), but have an impact on hardware cost, complexity, and compatibility with other equipment.

Racks (S-2): The idle power consumption of a server can be more than 50% of its peak power consumption [Takouna et al. 2011]. Moreover, "most servers consume between 70 and 85 percent of full operational power" [Emerson 2009], which certainly does not represent a proportional increase of energy consumption with respect to system output.

Consequently, "a facility operating at just 20 percent capacity may consume 80 percent of the energy as the same facility operating at 100 percent capacity" [Emerson 2009]. Additionally, this includes a huge energy waste by running servers idle without any useful output or with low utilization in the 10–50% utilization range, which is usually the case in typical data centers [Hoelzle and Barroso 2013]. Finally, racks containing components that are not used at all (e.g., graphics cards) also contribute to energy loss. Another source of energy loss is fans, which have typical efficiency of around 60% (i.e., around 40% of power is lost due to heat dissipation). Additionally, if fan speed is not well adjusted to server load and temperature, a significant amount of energy is wasted.

Components (S-3): The energy efficiency of server components drastically affects the overall efficiency of a server. A CPU take a bigger slice of the total energy consumption, which is more than a third of total server energy consumption. However, servers with large number of slower CPU cores can lead to lower utilization.

To mitigate all this energy loss and waste, a number of actions can be performed. These actions include:

- **L1.** Reduce the heat load of server components such as the CPU. This can be achieved by using more energy-efficient components and their architectures; for example, using slower, so-called wimpy CPU cores that are more power efficient [Mudge and Holzle 2010], as in the FAWN project [Andersen et al. 2009] where they utilize wimpy cores to build an energy-efficient key value storage system. Another recognized approach is limiting input energy to a specific component (S-1) or an entire rack (S-2), also referred to as power capping [Bhattacharya et al. 2012]. Similarly, in the case of the memory subsystem, performance can be adjusted (i.e., throughput is used to mitigate high temperatures) and thus avoid energy loss through heat [Lin et al. 2007]. Energy loss can also be reduced by using compact server configurations that exclude components that are not used (e.g., Google uses such an approach in building its data centers).
- **L2**Additional energy can be saved by reducing energyconsumed by supporting systems, such as cooling and power supplies inside theserver enclosure (S-3) and the servers themselves (S-2). For example, Google placesbackup batteries next to racks, thereby avoiding large UPS units that require theirown cooling systems [Wired 2013].
- **W1** Use components that can automatically scale their power consumption based on current load choosing the right processor architecture can also contribute to more efficient energy usage.
- **W2** Bigger cache size does not necessarily mean a lower miss rate. Therefore, choosing the right size cache can decrease energy waste Additionally, using cache subsystems for storage disks to reduce reads and writes from/to the disk and increase its idle time can also contribute to energysavings [Wang et al. 2008]. Such onboard controller caches can already be found on modern hardware.

E. Cloud Management System (CMS)

Managing and monitoring a cloud infrastructure with regard to energy efficiency and consumption has been identified as the main concern within a data center facility according to Emerson [2010]. Thus, the CMS plays an important role in trying to improve efficiency, increase utilization, and thus lower energy loss/waste within a data center.

The CMS domain includes the scheduler, monitoring system, virtualization technology, and all other software components responsible for managing physical and virtual machines within a cloud (e.g., OpenStack [OpenStack 2012] and Xen hypervisor [Citrix 2012]). A scheduler's main function is to deploy resources for fulfilling customer requests. Its supporting task is providing a monitoring system that gives additional information about allocated and available resources, such as utilization, QoS. Additionally, virtualization technology is used for better resource management and on-demand deployment and offers high scalability for cloud infrastructures. Based on this, the energy efficiency of the CMS can be examined through its component systems and includes both energy loss and waste, as shown in figure 6.

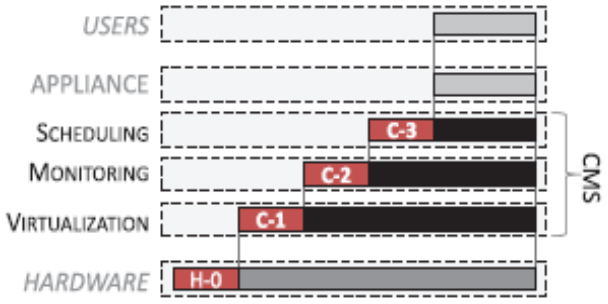

Figure 6 Losses and wastes of CMS domain.

Figure 6 also represents energy delivered to hardware equipment that was not fully utilized by the CMS domain (e.g., idle machines). Although H-0 can be directly related to the Hardware domain, it can also be minimized from the CMS perspective, for example, by consolidating underutilized machines and turning off idle ones [Feller et al. 2012].

Toreduce these wastes and loses, a number of actions can be taken which are as follows:

- **L1** During the development phase of the CMS, implement functions that can directly control hardware equipment because the CMS has "knowledge" of which resources are required and which are not (e.g., shutting down idle machines [Borgetto et al. 2012b]). The CMS can go beyond controlling only servers to expand its control to the network system or even to the cooling and power supply system's [Lago et al. 2011]. In this way, energy delivered to hardware equipment that is not utilized by the CMS (H-0) could be significantly reduced.

- **L2** the CMS should use lightweight supporting subsystems, such as monitoring (C-2) and virtualization (C-1) technologies, and avoid cumbersome systems that provide large numbers of functionalities that are not utilized by the cloud manager. This includes lightweight monitoring systems [Ma et al. 2012] and the selection of appropriate virtualization technology, namely, full- virtualization vs. para-virtualization or even microkernel architectures [Armand and Gien 2009].
- **W1**. Running the CMS supporting systems idle still consumes resources and therefore wastes energy (C-1 and C-2). For this reason, CMS subsystems should be implemented in a modular fashion, in which modules are loaded only when they are actually required (e.g., the monitoring agent that loads plugins for initialized metrics and removes them once they are no longer required [Mastelic et al. 2012]). This also includes minimizing resource consumption while running in idle mode (e.g., using lightweight hypervisors).
- **W2**. CMS system energy waste (C-3) can be avoided by optimizing the measuring not only its results, but also its tradeoffs for achieving those results (e.g., how much resources a single scheduling action takes and how many actions). This includes optimization of the scheduling algorithm and technology used for its implementation.

F. Appliance

The Appliance subdomain represents a part of the Software domain, which performsactual useful work for cloud users. In a perfect cloud infrastructure, only Appliances would be consuming resources and thus energy. From a provider's perspective, appliance efficiency is only considered for SaaS and PaaS applications because an applianceis then under control of the provider and thus part of the cloud computing infrastructure. On the other hand, for lower level services (e.g., IaaS), an appliance is deployed by a user, thus the user is responsible for its efficiency. This scenario falls under the User domain perspective. To date, software designers usually look at the quantity and proportionality ofperformance given the resource utilization. Now, to ensure energy efficiency, software designers also need to consider the quantity and proportionality of resource utilization given the performance.

The Appliance has a relatively smaller impact on the overall energy consumption than some other elements of the cloud infrastructure such as servers. On the other hand, appliances are responsible for the useful work, which is ultimately delivered to users. Hence, to adequately assess and manage the energy efficiency of the cloud, appliances must be taken into consideration. Three subsystems can be recognized in the appliance: an application that is used by the end user and that performs a main task of the appliance, a runtime environment required for running the application, and an operating system, which serves as a bridge between the physical or virtual machine and the software running on top of it.

The energy efficiency of appliances affects both energy loss and waste as shown in figure 7.

Figure 7 Losses and wastes of appliance domain

In addition to applications, runtime environment and operating system from the Appliance perspective, energy spent for running the CMS (C-0) is considered as entirely lost because it performs supporting tasks rather than the main task of the appliance. A number of actions can be taken to reduce these energy losses and wastes. These are as follows:

- **L1**Proper implementation of cloud appliances can help to reduce energy losses. This can be done during the development phase by optimizing the implementation, as well as by using lightweight programming languages and only required libraries (A-2). A fine-grained monitoring and estimation of power usage can be used in order to identify processes responsible for high energy consumption.
- **L2**. Although the Appliance sub domain represents IT software equipment, it can still have supporting systems that cause energy losses by creating resource consumption overhead (e.g., a small application running on a heavy operating system) while using only a small percentage of its functions (A-3). Therefore, energy losses can be reduced by proper implementation of applications and selection of an appropriate underlying technology. This should also include the use of reduced and customized operating systems and runtime environments. Similarly, as in the case of L1, precise information about the parts of software responsible for high energy consumption must be identified to apply appropriate optimization.
- **W1**. Optimization of the appliance can reduce energy consumption by decreasing its resource usage or increasing its performance. To achieve low wastes, decisions should take into account available hardware so that the number of threads/processes or internal load balancing areoptimized with energy efficiency in mind. Additionally, appliances need to be highly scalable in order to fully utilizeavailable resources.
- **W2** Minimizing the unnecessary use of the appliance depends on the way users access it. Any smart functions applied must avoid breaking SLAs set with users. However, even with these constraints, a number of actions can be taken to reduce useless energy consumption. These techniques can include serving requests in batches,

reducing the numbers of backups and checkpoints, limiting the number of service instances or threads, and adjusting the frequency of monitoring, polling, caching, and indexing. Overheads can be also related to the relevant functionality of appliances (e.g., security and resilience). Hence, applying most of these techniques requires finding a tradeoff between energy efficiency and other key aspect of appliances, such as performance, resilience, and security.

III. CONCLUSION

In this article, the energy efficiency of a data center's ICT equipment is analyzed, including the hardware and software that drives the cloud computing. The analysis showed that many standard energy efficiency techniques do not work forcloud computing environments out of the box; rather, they have to be at least adaptedor even designed from the scratch. This is due to the stratification of the cloud computing infrastructure, which comprises systems and components from different researchareas, such as power supply, cooling, computing, and more. Optimizing these systemsseparately does improve the energy efficiency of the entire system; however, applyingshared energy-efficiency techniques to multiple systems or their components can significantly improve energy efficiency if the techniques are aware of their interactionsand dependencies.

REFERENCES

[1] Advisory Group. 2008. *ICT for Energy Efficiency*. Technical Report. DG-Information Society and Media, European Commission. Retrieved from http://ec.europa.eu/information_society/activities/sustainable growth/docs/consultations/advisory_group_reports/ad-hoc_advisory_group_report.pdf.

[2] Dennis Abts, Michael R. Marty, Philip M. Wells, Peter Klausler, and Hong Liu. 2010. Energy proportional datacenter networks. *SIGARCH Computer Architecture News* 38, 3 (June 2010), 338–347.

[3] U.S. EIA. 2013. *International Energy Outlook 2013 with Projections to 2040*. Technical Report. U.S. Energy Information Administration, Office of Energy Analysis, U.S. Department of Energy. Retrieved fromhttp://www.eia.gov/forecasts/ieo/pdf/0484(2013).pdf.

[4] Emerson. 2008. *Data Center Users Group Special Report: Inside the Data Center 2008 and Beyond*. Technical Report. Data Center Users' Group, Emerson Electric Co., Emerson Network Power, Ohio.

[5] Emerson. 2009. *Energy Logic: Reducing Data Center Energy Consumption by Creating Savings That Cascade across Systems*. Technical Report. Emerson Network Power, Emerson Electric Co., Emerson Network Power, Ohio.

[6] Emerson. 2010. *Data Center Users Group Special Report: Growing Concerns over Data Center Infrastructure Monitoring and Management*. Technical Report. Data Center Users' Group, Emerson Electric Co.,Emerson Network Power, Ohio.

[7] Enerdata. 2014. Global Energy Statistical Yearbook 2014. Retrieved from http://yearbook.enerdata.net/.EU. 2008. *Addressing the Challenge of Energy Efficiency through Information and Communication Technologies*. Technical Report. European Commission, European Communities.

[8] R. Beik. 2012. Green cloud computing: An energy-aware layer in software architecture. In *Proceedings of the 2012 Spring Congress on Engineering and Technology (S-CET'12)*.1–4.DOI:http://dx.doi.org/10.1109/SCET.2012.6341950

[9] Anton Beloglazov, RajkumarBuyya, Young Choon Lee, and Albert Y. Zomaya. 2011. A taxonomy and survey of energy-efficient data centers and cloud computing systems. *Advances in Computers* 82 (2011), 47–111.

[10] D. Dharwar, S. S. Bhat, V. Srinivasan, D. Sarma, and P. K. Banerjee. 2012. Approaches towards energy-efficiency in the cloud for emerging markets. In *Proceedings of the 2012 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM'12)*.1–6.DOI:http://dx.doi.org/10.1109/CCEM.2012.6354599

[11] EARTH. 2011. Homepage. Retrieved from https://www.ict-earth.eu/.

[12] ECONET. 2013. Homepage. Retrieved from http://www.econet-project.eu/.

[13] Anton Beloglazov and RajkumarBuyya. 2010. Energy efficient allocation of virtual machines in cloud data centers. In *Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster,Cloud and Grid Computing (CCGRID'10)*. IEEE Computer Society, Washington, DC, 577–578. DOI:http://dx.doi.org/10.1109/CCGRID.2010.45

[14] I. Foster, Yong Zhao, I. Raicu, and Shiyong Lu. 2008. Cloud computing and grid computing 360-degree compared. In *Proceedings of the 2008 Grid Computing Environments Workshop (GCE'08)*.1–10.

[15] Chang Ge, Zhili Sun, and NingWang. 2013. A survey of power-saving techniques on data centers and content delivery networks. *IEEE Communications Surveys Tutorials* 15, 3 (Third 2013), 1334–1354.

