From Reducing Energy Consumption to Reducing CO_2 Emissions: The ECO_2Clouds Approach

Usman Wajid[1] and Pierluigi Plebani[2]

[1]University of Manchester, Manchester, U.K.
[2]Politecnico di Milano, Milano, Italy
usman.wajid@manchester.ac.uk
{pierluigi.plebani, barbara.pernici}@polimi.it

Abstract. While the topic of energy efficiency (in cloud and large scale datacenters) has been attracting significant interest from academia and industry owing to mainly economic and somewhat environmental implications, reducing CO_2 emissions is often side-lined as a consequent benefit. This paper presents a more direct approach adopted in the ECO_2Clouds project for reducing CO_2 emissions in cloud and scale datacenters. The ECO_2Clouds approach relies on (a) definition of key metrics to enable quantification of CO_2 footprint at application and infrastructure level, (b) CO_2 aware deployment strategies, and (c) adaptive management of workloads to constantly minimize the CO_2 footprint of applications as well as underlying resources/infrastructure.

1 Introduction

With rapid proliferation of large scale datacenters and cloud facilities, tackling the issue of energy consumption has emerged as one of the critical research challenges to be dealt with. In response, a plethora of approaches have proposed ways to achieve energy efficiency and harvest associated benefits that mostly translate into economic advantages.

Besides energy efficiency, a related stream of research focuses on evaluating and reducing the environmental impact of cloud computing by means of minimizing CO_2 emissions from cloud and datacenter facilities. The issue of CO_2 emissions resulting from vast energy consumption is a growing concern that is generating economic, social and political pressure. This pressure from different sectors of society is likely to result in regulations as well as influence consumer selection criteria for outsourcing. Existing work in this area mostly ranks CO_2 reduction or environmental impact as a consequent benefit of achieving energy efficiency. However, CO_2 derived from the different energy sources is often not addressed directly, and relatively little attention is diverted towards effective utilization and optimization of available energy sources. Hence there are opportunities for addressing environmental implications by adopting a more direct approach for reducing CO_2 emissions.

In this background, the European Commission funded ECO_2Clouds (Experimental Awareness of CO_2 in Federated Cloud Sourcing) project aims to addressing the environmental impact of cloud computing by not only developing methods for quantification of energy consumption and CO_2 emissions at different levels of cloud computing

Pernici B. and Plebani P.
From Reducing Energy Consumption to Reducing CO2 Emissions - The ECO2Clouds Approach.
DOI: 10.5220/0006183200610070
In *European Project Space on Information and Communication Systems (EPS Barcelona 2014)*, pages 61-70
ISBN: 978-989-758-034-5

(e.g. testbed, host, virtual machine and application level) but also allowing for proper consideration of gathered data at application deployment and execution lifecycle. ECO_2Clouds develops and advocates a reinforcement learning model that relies on monitoring of ecological parameters in cloud computing and using this information to improve the way applications are designed and how infrastructure responds to changes occurring at physical hosts, running applications, down to energy mix consumed.

The ECO_2Clouds solution is particularly designed for federated clouds that (a) offer opportunities of utilizing different cloud resources, and (b) enable minimizing CO_2 footprint of applications by considering energy mix of different testbeds in the applications deployment decision making model. In this respect, the availability of different and geographically separated infrastructure or testbeds allow distribution of applications on eco-friendly resources, where permitted by testbed and availability of suitable resources required by applications.

The ECO_2Clouds is described in Section 2. Section 3 describes how the quantification of environmental impact of cloud computing is carried out in ECO_2Clouds. Section 4 presents the deployment and adaptation techniques devised in ECO_2Clouds in order to minimize the environmental impact of cloud computing. Section 5 presents discussion of some preliminary results of ECO_2Clouds solution. Section 6 concludes paper with directions for future work.

2 An Overview of ECO_2Clouds Approach

ECO_2Clouds adopts an iterative development approach involving three key phases namely Measure, Create and Test, shown in Fig. 1.

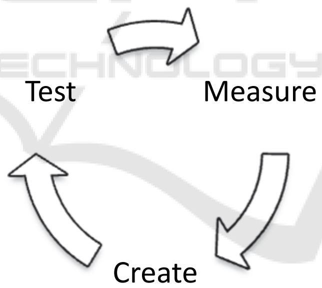

Fig. 1. Three phases of ECO_2Clouds approach.

The *Measure* phase focuses on quantification of energy consumption and environmental impact of cloud computing, the *Create* phase develops techniques and software artefacts to help realize awareness of the environmental impact of cloud computing by making improvements in the reduction of energy consumption and CO_2 emissions, and the *Test* phase tests the outcome of previous two phases on an existing FIRE (www.ict-fire.eu) facility known as BonFIRE (www.bonfire-project.eu).

The three phased approach addresses important research questions such as quantification of environmental impact of cloud computing, enacting deployment and runtime adaptation actions that can decrease the energy consumption and CO_2 footprint of cloud computing and considering environmental implications in the design and subsequent execution of cloud applications.

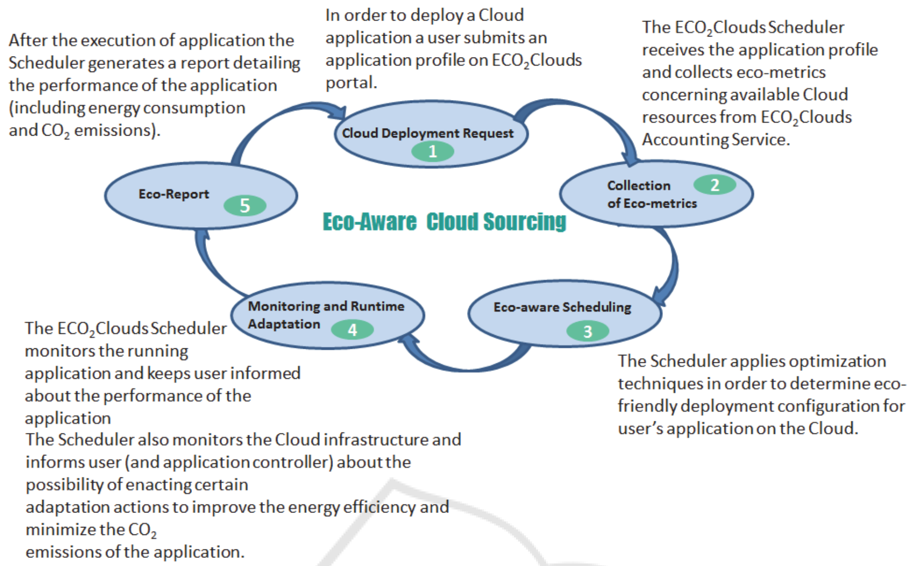

Fig. 2. Implementation of ECO_2Clouds approach.

Fig. 2 provides an overview of implementation and execution aspects of ECO_2Clouds approach where the workflow is triggered by an application deployment request. A deployment request typically entails the availability of certain number of virtual machines with certain characteristics (e.g. CPU and Memory requirements). The deployment of virtual machines is determined by the ECO_2Clouds Scheduler, a key decision making component in the system. The subsequent step involves the collection of eco-metrics by Scheduler from different levels of the cloud infrastructure. Eco-metrics is a term used in ECO_2Clouds to refer to parameters that provide information concerning energy consumption and CO_2 footprint at different levels of cloud infrastructure. The information provided by eco-metrics is used in the eco-aware scheduling of applications on the federated cloud resources, as shown in step 3. Once deployed the applications are constantly monitored along with the way the underlying infrastructure behaves e.g. how spikes in resource utilization, deployment of new applications and termination of already running applications provide opportunities for further improvements in energy consumption and CO_2 footprint of available resources. The monitoring information is taken into account while deciding and triggering certain adaptation actions that can reduce the energy consumption and CO_2 footprint of running applications as well as the underlying infrastructure. After the execution of applications, the detailed monitoring information is compiled in an eco-report for user.

The report informs the user (e.g. application designer and developer) about total energy consumption and CO_2 footprint of their applications. Based on the reinforcement learning model of ECO_2Clouds the user is encouraged to use the information from eco-report to improve the design and execution aspects of their applications in a bid to make them more energy efficient and eco-friendly. Furthermore, it is envisioned that the awareness about energy consumption and CO_2 footprint can enable the application designers and developers to better tune their applications e.g. by more effectively utilizing the available resources or by choosing adequate resources for the deployment and execution of application.

3 Quantification of Environmental Impact

As the final goal of the ECO_2Clouds project is to reduce the environmental impact of applications running on a federated cloud infrastructure, it is crucial to properly quantify and monitor such an impact.

The energy consumed by an application is usually taken as a primary metric: the more the application consumes, the more impact the application has on the environment 2. Although calculating the energy consumption is important, this captures only a fraction of the real impact. Indeed, it is also important to evaluate which are the sources that provide such energy: Nuclear plants? Coal plants? Renewable sources? As a consequence, CO_2 emissions are usually taken as a good metric for evaluating the real impact, so that, an application that consumes an amount of energy coming from a renewable source is preferable to an application that consumes the same amount of energy coming from a coal plant.

It is worth noting that, although considering the CO_2 emissions during the execution is reasonable, the computation of CO_2 emission could be tricky. For instance, renewable sources are correctly considered as zero-emission sources. This is true if we consider the emission during the energy provisioning. On the contrary, if we consider the complete life cycle of renewable sources as hydro-power plants or wind farms, the CO_2 emission cannot be considered zero as some energy coming from CO_2 emitting sources is required to build, and it will be required to dismiss, the plants. This issue becomes more important when considering nuclear plants. Indeed, CO_2 emissions of this kind of plants are almost zero, by contrast the energy required to build for the decommission of a nuclear plant is incredibly high 3. As at this stage it is very difficult to really compute this holistic CO_2 emission, the ECO_2Clouds approach quantifies only the CO_2 emitted when producing the required energy. Anyway, information about the energy sources are considered in order to increase the awareness of the user about the kinds of energy sources involved when the application required are used.

On this basis, the ECO_2Clouds project proposes a layered approach to monitor the environmental impact of the applications, the virtual machines, and the sites as shown in Fig. 3. For each of the layers, a set of metrics to be monitored has been identified 1. Each of these sets of metrics captures not only aspects about the environmental impact, but also the performances of the system. This allows our solution to mediate between improvements with respect the environmental sustainability and the performances.

More specifically, starting from the bottom, the infrastructure layer includes metrics that measure the behavior of the single physical machines and the entire site. Especially for the energy metrics, this layer is fundamental, as it is the place in which the power can be physically measured using PDU (Power Distribution Units). Along

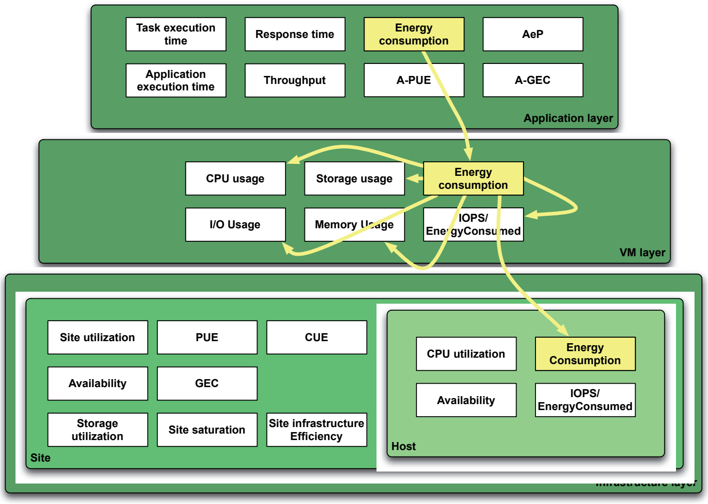

Fig. 3. Layered architecture in $ECO_2Clouds$.

with the classical eco-metrics as PUE (Power Usage Effectiveness), the environmental impact is taken into account by measuring the GEC (Green Energy Efficiency) 4: i.e., the percentage of the power coming from renewable source. Yet, availability of the resources (CPUs, memory, and storage) is also measured to monitor the possibility to deploy application on a site.

Moving to the next layer, the VM layer focuses on the metrics able to evaluate the status of the VMs running on the physical host. At this layer, most of the metrics are derived from the metrics measured at the infrastructure layer. Focusing on the environmental impact, here the energy consumed to run a VM is computed by considering the amount of resources reserved and really used by the VM in order to properly subdivide the energy consumed by the physical host among the VMs running on it.

At the top-most layer, there are the metrics that evaluate the applications running on the VMs. Assuming that an application can be distributed among several VMs and these VMs can live on different sites, the metrics included at this layer, compute the environmental impact as well as the performances starting from the metrics at the lower levels. For instance, the energy consumption of the application is obtained considering the execution time and the energy consumed by all the VMs required by the application. At this level, it is worth noting that metrics inspired by the usual metrics included in the infrastructure layer are also included. For instance, the A-PUE (Application PUE) has been inspired by the classical PUE. As the latter is defined as the quantity of power used by the site divided by the power really used by the IT devices, the A-PUE is defined as the quantity of power used by all the VMs involved in the application divided by the power really used for running the application. This parameter has been considered as, usually, when running an application on a VM several other processes are running as well. Some of them are required (i.e., operating system processes), some other starts automatically when the VM boots but they are not really needed (i.e., a mysql deamon or an http server even if the application does not need them). Goal of this metric is to make the user aware of this fraction of power that is not consumed specifically for the application. More details on the set of metrics and the facilities involved to monitor them can be found in 1.

4 ECO-Aware Application Deployment and Adaptation

At primitive level the ECO_2Clouds solution is driven by the Scheduler, a central entity or service responsible for determining eco-friendly deployment of application and later runtime adaptation of deployment configurations based on the availability of eco-metrics.

4.1 ECO-Aware Deployment

The ECO_2Clouds Scheduler implements different application deployment policies and allows system administrator to switch between the policies using a simple REST interfaces. The deployment policies that can be characterized as eco-aware heuristics perform decision making at two different level i.e. at testbed and physical host level.

Keeping in view the federated cloud infrastructure, once a deployment request is received, the first Scheduling step is to select a suitable testbed. This step is realized by the switching the Scheduler in one of the following modes:

Individual Deployment Mode performs selection of a testbed for each individual VM in the deployment request. In this respect, the VMs in a single deployment request (representing a distributed application) may be deployed on different testbeds, for example after the allocation of a single VM the suitability of a testbed may change, thus the next VM in the same deployment request may be allocated to another testbed that fits the suitability criteria.

Bulk Deployment Mode performs selection of testbed for all VMs in a particular deployment request. In this mode, all VMs (belonging to an application) will be deployed on a single suitable testbed.

In the above two modes, the suitability of a testbed is determined by a combination of multi-criteria optimization and load balancing functions. The multi-criteria function takes into account a number of eco-metrics at testbed level with particular emphasis on energy consumption, CO_2 footprint (determined from energy mix consumed by the geographically distributed testbeds), PUE and GEC. Once a testbed is selected, the following deployment policies can be used to determine physical host level deployment configuration of new VMs/applications.

Max-Utilization or Task Consolidation tries to maximize the utilization of individual physical hosts by deploying VMs on minimum number of hosts (e.g. hosts with most CPU utilization and highest energy consumption) while keeping as many hosts idle as possible. The idle hosts can be switched off or switched to hibernate mode provided adequate support by the underlying cloud infrastructure.

Min-Utilization or Task Dispersal, which tries to minimize the utilization of individual hosts by deploying VMs on least used hosts (e.g. hosts with most free-CPU and lowest energy consumption) in order to balance the workload across available hosts.

The combination of deployment modes and policies help in determining an eco-ware deployment configuration for cloud applications. The experimental evaluation of the above deployment modes and policies is yet to determine their effectiveness over each other, however the consideration of environmental impact (by means of considering energy consumption as well as CO_2 emission derived from energy sources) is a step forward in realizing ecologically aware cloud computing.

4.2 Eco-Aware Adaptation

In ECO_2Clouds adaptation is referred to a process of change that makes applications and underlying cloud infrastructure more energy efficient and environmentally friendly. In ECO_2Clouds adaptation occurs at two different levels:

- Infrastructural level adaptation: the ECO_2Clouds Scheduler drives the adaptation by properly managing the VMs.
- Application level adaptation: the Application Controller, an application dependent module, drives the adaptation by distributing the workload among the VMs or by requesting the allocation and the release of VMs.

4.2.1 Adaptation at Infrastructure Level

In ECO_2Clouds the deployment decision making entity i.e. Scheduler only works at the VM level and has no control over what go on inside a VM (or an application). In this respect, the adaptation driven by Scheduler tries to achieve the minimum energy consumption and CO_2 footprint at the infrastructure level without interfering directly with the running applications. The overall objectives of adaptation at infrastructure level are:

- Efficient utilization of available resources in order to keep the energy consumption and CO_2 footprint of cloud infrastructure to a minimal level
- Fulfilment of requests from running applications

The above objectives are achieved by enacting the following actions at infrastructural level:

- Supporting applications in starting, freezing and restarting VMs based on applications' internal adaptation criteria
- On request (from running applications) allocation and termination of VMs
- Improve the detail of observation of one metric (e.g. change granularity of metric by varying the frequency of sampling)
- During the runtime of applications, informing users about the potential of moving certain VMs on more energy efficient and/or CO_2 friendly resources such as testbeds and physical hosts.

4.2.2 Adaptation at Application Level

Adaptation at infrastructure level is based on information regarding the status of the physical hosts and virtual machines. Another source of useful information that needs to be considered is at application level. For example, at infrastructure level, the Scheduler notices that a VM has a low CPU load and it might need to ask, the infrastructure layer, to move the VM to another physical host so the one in which the VM is currently running can be switched off. Although this choice sounds reasonable, the Scheduler cannot be aware of the fact that the status of low CPU load is only temporary and in no more than 10 minutes the CPU load will drastically increase.

To realize application level adaptation ECO_2Clouds prescribes that cloud applications implement an internal management module or Application Controller (AC). AC is envisioned as a complementary component responsible for providing a set of facilities for accessing the eco-metrics related to the application and enabling the adaptation at application level. The application designers can implement the AC based on the nature of their applications and best possible adaptation scenarios.

The role of the Application Controller is to read the application level metrics and to find the best configuration in order to optimize the energy efficiency, reduce CO_2 footprint and avoid wasting of resources. This can be done by properly distribute the workload among the several VMs reserved for an application, changing the state of running VMs based on application specific conditions, or to ask to change the number of VMs in case they are either too much or not enough. In this way, the application controller tries to avoid as much as possible, situations in which energy is wasted and resources are underused.

5 Discussion

The ECO_2Clouds solution is implemented as an extension of the BonFIRE platform. This BonFIRE platform provides a federated cloud infrastructure for deploying and running VMs of different sizes (e.g., according to the usual nomenclature: small, medium and large). The platform also provides a monitoring system based on Zabbix[1] to evaluate the metrics at infrastructure and VM level. In ECO_2Clouds, this set of metrics has been extended to cover the requirements introduced in Section 3.

When running an application, ECO_2Clouds platform is able to monitor and collect the values of all the metrics presented before. Some of the values can be viewed in Fig. **4**. Here two application level metrics, i.e., *Application Throughput* and the *Application Energy* are compared to identify possible energy wasting. Indeed, the user is in charge of improving the application and can figure out when there is a discrepancy between the trend of the energy consumed and the effective work done. In this case, the evaluation can be only qualitative as these two metrics are not directly comparable.

In some other cases, the metrics can be compared and some adaptation decisions can be taken. For instance, Fig. **5** shows the trend of the power consumed by the two VMs used by an application. In this case, we have again some wastage as the

[1] http://www.zabbix.com/

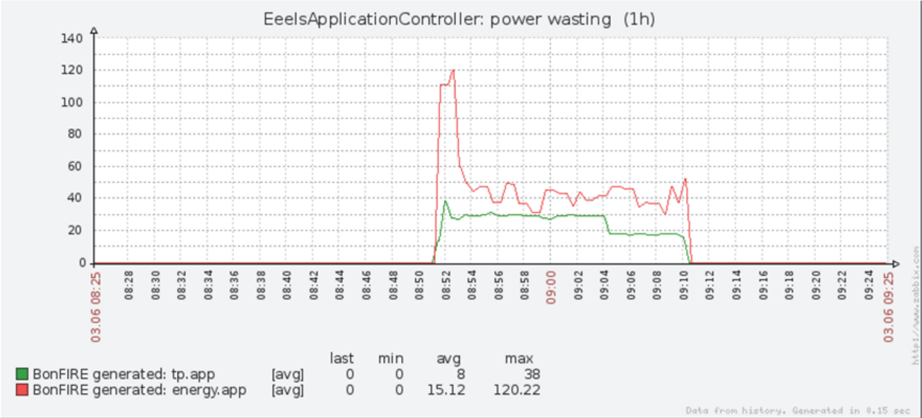

Fig. 4. Throughput against energy chart (best viewed in color)**.**

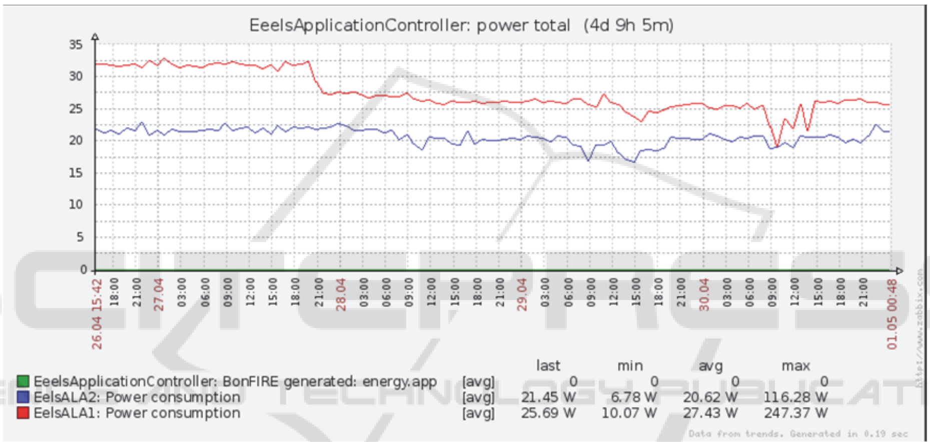

Fig. 5. VMs and Application power chart (best viewed in color)**.**

power of the application is constantly zero (as shown in the legend). As evident in the figure, during in the reported period the VMs are running but no useful work for the application is done since the application parts deployed in the VM are not running. In such situations, when the VMs are running uselessly, the Application Controller can take an adaptation decision e.g. by requesting the Scheduler to change the state of the VMs or switch them off in order to conserve energy and make resources free for other applications. In both cases, the adaptation actions will directly contribute towards reducing the energy consumption and subsequent CO_2 footprint of applications.

6 Conclusion and Future Work

The ECO_2Clouds project adapts a more direct approach towards tackling environmental impact of cloud computing by considering not only energy consumption but also

CO_2 footprint in its cloud sourcing model. The emphasis on CO_2 footprint and environmental impact (primarily derived from energy sources) makes ECO_2Clouds approach stand out from many other efforts that only focus on energy efficiency and single site cloud infrastructures.

Current ECO_2Clouds is going through final stages of the development and implementation phase. The initial testing of some components has been already initiated. The next step will be to deploy system level tests, which will prove the effectiveness of the overall solution and its ability to achieve the ambitious objectives of reducing the energy consumption and CO_2 footprint of cloud computing, as set out in the project plan.

For further details and updates concerning ECO_2Clouds, the readers are referred to project website: www. eco2clouds.eu

Acknowledgements

This work has been partially supported by the ECO_2Clouds project (http://eco2clouds.eu) and has been partly funded by the European Commission's IST activity of the 7th Framework Program under contract number ICT- 318048. This work expresses the opinions of the authors and not necessarily those of the European Commission. The European Commission is not liable for any use that may be made of the information contained in this work. The authors thank all the participants in the project for common work and discussions.

References

1 C. Cappiello, S. Datre, M.G. Fugini, P. Melia, B. Pernici, P. Plebani, M. Gienger, A. Tenschert: Monitoring and Assessing Energy Consumption and CO2 Emissions in Cloud-Based Systems. SMC 2013: 127-132

2 Greenpeace: How Clean Is Your Cloud? 17th April 2012. Available on line: http://www.greenpeace.org/international/en/publications/Campaign-reports/Climate-Reports/How-Clean-is-Your-Cloud/

3 M. Lenzen, Life cycle energy and greenhouse gas emissions of nuclear energy: A review. Energy Conversion and Management 49: 2178–2199, 2008

4 The Green Grid Consortium: Harmonizing global metrics to measure data centre energy efficiency. Oct 2012. Available on line: http://iet.jrc.ec.europa.eu/energyefficiency/sites/energyefficiency/files/files/documents/ICT_CoC/harmonizing_global_metrics_for_data_center_energy_efficien cy_2012-10-02.pdf