Assessment of the Environmental Impact of Applications in Federated Clouds

Barbara Pernici[1] and Usman Wajid[2]

[1]*Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy*
[2]*School of Computer Science, University of Manchester, Manchester, U.K.*

Keywords: Environmental Impact of Applications, Federated Clouds, Adaptive Services, Eco-metrics.

Abstract: While in recent work the energy efficiency of cloud platforms has been emphasized and many approaches have been proposed to reduce the energy consumption in data centres, the evaluation of the environmental impact of applications running in cloud environments is still a research issue, as well as possible techniques to lower their environmental footprint. In the paper the approach taken in the ECO$_2$Clouds project towards assessment of environmental impact of applications and the evaluation of its potential reduction based on adaptive services is illustrated and discussed, including in the discussion the evaluation of alternative possible uses of eco-metrics towards reducing the environmental impact of applications.

1 INTRODUCTION

Energy consumption in large scale data centres and in clouds is a growing concern that has only recently received necessary attention. Several techniques have been developed to improve energy efficiency in data centre and cloud environments (e.g. Kolodziej et al., 2012); (Lindberg et al., 2012) and several metrics have been proposed to assess the obtained results (e.g. Kipp et al., 2012). In (Beloglazov et al., 2011) a comprehensive survey is provided on the existing approaches for achieving energy efficiency, with particular reference to dynamic power management approaches, both at the hardware and at the software level. At the software level, approaches range from management of workloads at the level of the single servers, to overall workload management at data centre and cloud level.

In cloud computing virtualization provides necessary basis for not only dynamic management of resources but also for making efficient use of available resources while guaranteeing quality of service. Lately research work in the area of cloud computing has been focusing on allocating virtual machines with the goal of improving energy efficiency in cloud infrastructure and reducing environmental impact of cloud computing [Kolodziej et al., 2012); (Lindberg et al., 2012); (Khosravi et al., 2013); (Wajid et al., 2013). On the other hand, to carry out assessment of energy efficiency techniques and approaches several metrics have been proposed (Kipp et al., 2012); (GreenPeace, 2012). In this respect, the GreenGrid Consortium has proposed a Data Center Maturity Model (DCMM) which, on the basis of computing, networking, infrastructure, and management variables enables users to determine the current performance of data centres and identify ways to achieve greater energy efficiency and sustainability (Singh et al., 2011). In this work strong emphasis is put on a better use of available resources based on parameters such as for instance the percentage of CPU usage, which, rather than being kept low in the average data centre, should be increased to reach 60% on the average to achieve desired outcome (envisioned by 2016).

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 the point of view of environmental impact, many proposals are based on deriving from the consumed energy the equivalent estimation of CO_2 emissions. However, in a (GreenPeace, 2012) report, several other factors are highlighted that can also contribute towards environmental impact of cloud computing.

In this respect, reporting of the Carbon Utilization Effectiveness (CUE) is one of the parameters which is starting to be considered by

some of the public cloud providers. Another contribution of GreenPeace report is the clean energy index, which considers the energy mix for generation of the utilized energy, siting, the adoption of energy efficiency techniques, and reuse of energy in datacentre or cloud environments.

Notwithstanding the many proposals there is not a single agreed or standard metric or set of metrics that clearly allow the assessment of the environmental impact of cloud computing. In particular, while the above-mentioned work focus on the IT infrastructure (and in some cases also the facility management, which is out of the scope of this work), there is no clear indication on how energy efficiency and environmental impact should be evaluated at the level of application deployed or running on cloud.

If we focus on the assessment of non-functional parameters at application level, adaptive approaches have been proposed ((e.g. in the S-Cube network, (Papazoglou et al., 2010)) as a way to guarantee quality of service in service-based systems where services are provided in a distributed way to a variety of users and to other service providers. In such environments, a major emphasis is given to the ability of assessing the context of execution, in order to base adaptation on the available knowledge of the context. In (Bucchiarone et al., 2012) such an approach has been particularly studied for composed services, where service is provided through a combination of services to clients requesting it with agreed Quality of Service constraints.

In this paper we describe how in the EU funded ECO$_2$Clouds project the assessment of environmental impact of cloud applications is performed, and by virtue of this assessment we evaluate the potential benefit of an adaptive approach that we have developed and implemented in the project. The focus being on the IT components of federated clouds and efficiently running applications in the cloud, considering also the dynamically varying energy mix in the different sites, we show how such an approach could bring more than an order of magnitude of improvement in the environmental impact of cloud applications.

In addition to the discussion about the potential of adaptive and context-aware approaches, the goal of the paper is to put forward a discussion about the possible uses of indicators, not limited to techniques for automatically reducing absolute values of energy consumption and environmental impact.

In this position paper, first, in Section 2, we briefly introduce ECO$_2$Clouds project and discuss the assessment of environmental impact of applications as carried out in ECO$_2$Clouds and possible approaches for evaluating the assessment. This is followed by a brief description of adaptation techniques used in the ECO$_2$Clouds project for reducing the environmental impact.

In Section 3 we evaluate the potential of using adaptation techniques on experimental data while discussing possible options. Finally, in Section 4 we discuss possible approaches for assessing the environmental impact of applications and their potential for improvement.

The paper ends with summary remarks and acknowledgements.

2 ENERGY EFFICIENCY AND ENVIRONMENTAL IMPACT IN ECO$_2$CLOUDS

The ECO$_2$Clouds project (http://eco2clouds.eu) investigates strategies and develops mechanisms for environmentally aware cloud sourcing with the aim to reduce energy consumption and CO$_2$ emissions of cloud applications as well as of the underlying infrastructure.

In ECO$_2$Clouds, assessment of environmental impact of cloud computing is carried out by defining a set of layered metrics. The evaluation of the metrics is performed within an adaptive and context-aware application deployment and management approach discussed in the subsequent sections.

2.1 Metrics and Assessment

In ECO$_2$Clouds, three levels for evaluation of environmental impact are considered: physical, virtualization, and application level. While it is not the goal of the paper to provide details on all metrics concerning each of the three levels (described in detail in ECO$_2$Clouds D3.1, 2013), we focus here only on the evaluation of energy and CO$_2$ emissions related parameters. In ECO$_2$Clouds assessment and improvements concerning these parameters relies on the management of IT component of the cloud infrastructure, and in particular on the computing components.

In ECO$_2$Clouds energy is evaluated at physical level of cloud infrastructure, based on measurements from PDU (Power Distribution Units), for each site of the federated cloud. From these measures, energy consumption is evaluated for each server or host of each cloud site. From measurements at host level, considering the running virtual machines and their

CPU usage, the energy consumption is mapped and evaluated for each virtual machine.

A third level introduced in $ECO_2Clouds$ concerns the actual use of VM by applications. We assume that one or more applications (or service) can be executed on a VM, and, based on their execution parameters, such as start and termination times and CPU usage for each service, power level and energy consumption are evaluated for each running service.

$ECO_2Clouds$ utilises a federated cloud infrastructure composed of three geographically distributed cloud sites located in Germany, France, and UK (details of the characteristics of the sites are provided in the $ECO_2Clouds$ web site), where the energy mix (i.e. proportion of coal, nuclear, hydro etc.) is known for each site. Such energy mix can vary based on the time of the day and of the year. Using the energy mix information alongside energy consumption measure the CO_2e (Carbon dioxide equivalent) emissions can be evaluated, for all three levels, i.e., physical, VM, and applications as a measure for environmental impact of the underlying cloud infrastructure and applications. In this way, it is possible to give to users a rather precise evaluation of the environmental impact of their cloud applications based on above-mentioned data.

2.2 Adaptation Actions

Based on the defined metrics, their evaluation can be the basis for establishing an adaptive control system for reducing the environmental impact of cloud computing.

As mentioned above, several approaches have been proposed in the existing literature for improving energy efficiency in the cloud (for example, as discussed in (Beloglazov et al., 2011) and (Lindberg et al., 2012).

In $ECO_2Clouds$, the problem of energy consumption and CO_2 emissions is addressed at the software level, assuming that optimization at the server and operating system level is not under control of the cloud platform or the users.

In this respect, actions that can be applied within the context of $ECO_2Clouds$ are at two levels i.e. VM level and application level. This limitation is primarily based on the use of BonFIRE platform (BonFIRE, 2013) and its functionalities that are made available for usage and experimentation in the $ECO_2Clouds$ project.

Further actions could be envisioned in other platforms; however, we briefly mention here the principal two types of actions under consideration, as a basis for the following discussion:

- VM level: here the main goal is scheduling of VMs on available cloud resources with a focus on consolidation, to be able to save energy by possibly switching off unused machines, and freeing less used computing resources for other uses.
- Application level: actions inspired by dynamic service composition approaches, VM reconfiguration and usage management of long running applications (such as the ones described in (Cappiello et al., 2013)), starting, freezing, reactivating, and terminating one or multiple VMs on which the application is running based on the applications' internal and execution context information.

3 POTENTIAL OF ADAPTATION

A first approach for a general evaluation of the environmental impact is to evaluate the total energy consumption and consequent CO_2 emissions over a given period of time.

In $ECO_2Clouds$, the Zabbix monitoring solution used in BonFIRE has been extended with the defined eco-metrics, since each cloud site in the BonFIRE federated cloud already provides a Zabbix server which uses Zabbix client agents to capture the cloud resource utilization data at host and VM levels. These Zabbix agents can be used to retrieve energy consumption and CO_2 related data for two of these layers: physical and virtual resources (Tenschert and Gienger, 2013).

Application level metrics are evaluated within the application deployment environment. The general architecture is shown in Fig 1. Three cloud sites are considered as testbeds in $ECO_2Clouds$.

An example of display of power metrics is shown in Figure. 2. In addition to the graphical interface and exports in spreadsheets, a REST interface for interacting with the application level control environment has been developed.

In this section we base our discussion on an analysis from the experimental data, starting from the assumption that in the beginning there is no VM management.

For an evaluation of the potential of considering the execution context of applications for their dynamic management, we consider for a rough evaluation an average CO_2e conversion factor for each given site. A further improvement could be gained considering also daily and seasonal variations in the energy mix, but we do not consider this in this paper to simplify the presentation, with the goal of

giving an approximate estimation of the possible gains. However, the energy composition is actually available as a monitored parameter within Zabbix, together with the other eco-metrics defined in $ECO_2Clouds$ project.

We assume to run applications initially on 10 separate physical servers, running with an average power of 133 W. Due to the different energy mix in the three countries, as reported in ($ECO_2Clouds$, D3.1, 2013) we can assume an average value of 85 gCO2e/kWh for France, 706 gCO2e/kWh for Germany, and 567.17 gCO2e/kWh for UK.
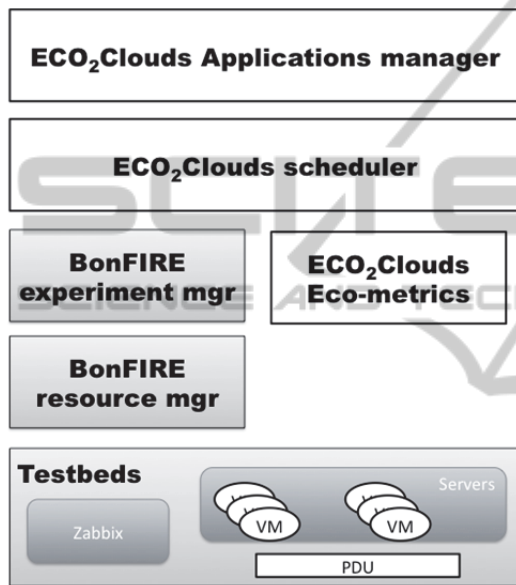

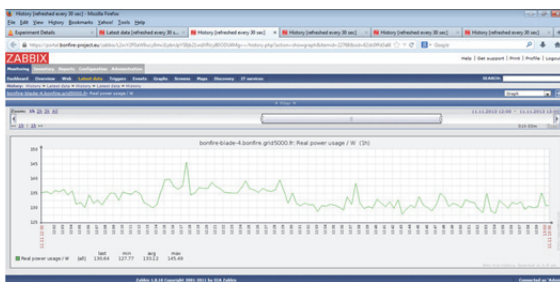Figure 1: $ECO_2Clouds$ architecture.

Figure 2: Monitored eco-metric (courtesy of INRIA).

Due to the application of VM consolidation techniques, based on the target goals of DCMM (GreenGrid, 2011), we can assume that the load of the servers can raise from 20% to 60%, owing to activating and deactivating the VMs as needs arise.

However, when running VMs in a BonFIRE cloud, we have also to assume that an experiment (i.e. the allocation of VMs to a user according to his requests) is immediately started, and in addition,

since a duration for the experiment has to be specified in advance, it is likely that the user will have to ask for more resources and for a longer time than needed because in case the application is not completed at the end of the requested period, the experiment will be stopped (and in this case will have to be started again).

From application level it is difficult to assess the potential reduction of environmental impact that can be gained from the ability to manage VMs, thus there are opportunities for user and the application to better control and manage their allocated resources. However, this depends on the initial configuration of the experiment, as well as from the type of the applications. For instance, in our experimentation first simulations show a potential impact from 10 to 50%.

	Wh per year	max CO2e	min CO2e	max energy reduction	max CO2e reduction
initial configuration	11650800	8225464.8	990318	-	0.88
server consolidation	3883600	2741821.6	330106	0.67	0.96
VM application level mgr	1941800	1370910.8	165053	0.83	0.98

Figure 3: Adaptivity analysis vs initial configuration.

In Fig. 3 we report a first evaluation of the potential reduction, with respect to the given initial situation. In the figure it is clear that in cases such as the one considered in the cloud federation, running in three different countries, the impact of the location is preponderant, and, combined with dynamic resource management, the potential gain could be on the order of two orders of magnitude. However, some additional considerations are necessary and will be discussed in the next section.

4 DISCUSSION ON EVALUATING THE ENVIRONMENTAL IMPACT OF CLOUD APPLICATIONS

From the reported example, it becomes clear that the simple definition of eco-metrics is not sufficient for an assessment.

First of all, a coarse evaluation based on total energy and CO_2e emissions can be misleading in some cases. In fact, it might lead to allocating the load only on a few of the available servers, and only in a single country when the energy mix is very different in the federated cloud.

In addition, it has to be considered that reducing power does not always imply reducing energy

consumption (see, for instance the discussion in (Chen et al., 2011)). In fact, applications might take longer to execute, so the reduction in energy consumption is not necessarily proportional to the reduction of power, and in some situations might even result in an increase of power consumption.

Alternative approaches can also be considered, based on the available metrics, to take into consideration other factors. One aspect is that considering only energy and CO_2e could result in violations of other parameters, for instance Quality of Service. To avoid this, assessment of environmental impact should be based on the evaluation of many factors. As in service-oriented systems, this approach can result either in considering a set of thresholds to be satisfied, or in defining a utility function for a global assessment of the system. In this case, other factors could be considered, such as the importance given by the users to some quality parameters or for a given energy mix (e.g., excluding coal or nuclear sources). Since the evaluation of importance could vary based on the user preferences, or even based on the specific nature of each single application, the user should be enabled to establish his own preferences for making an evaluation.

Other factors, even if not relevant at the moment within the BonFIRE platform as it is currently provided, could be linked in the future to variable costs of the provided service under different conditions such as the use of specific infrastructure at specific time, under specific QoS parameters and under specific energy consumption or CO_2 footprint related constraints.

These considerations would result in different evaluations when considering the user views and cloud provider view, which would give a different emphasis on different parameters. A comparison in such cases might be difficult, as well as an assessment of the final result.

A completely different approach could be based on management considerations, such as avoiding critical events (as proposed, for green IS, in (Reiter et al., 2013)). Critical events in cloud computing, when focusing on energy and environmental impact, could be the need to avoid reaching power limits in a data centre, which usually results in limited service or in additional costs for energy provisioning, or using on-demand resources in one of the sites of the federated cloud when other resources are available instead in other sites.

Finally, giving the users information about the energy consumption and related monitored information could raise awareness of ecological implications of their computing requirements, which in turn could encourage modification in user behaviour. For instance, if a user needs to report CO_2e emissions for his applications, and he sees a potential for reduction while running the application in a given site or in a given time of the day, he might schedule to run his applications, if not time critical, taking also these factors into consideration. In other cases, the user might realise that his use of allocated VMs is not optimal (in terms of energy consumption or CO_2e emissions) in a specific deployment configuration and that different configurations might reduce wastage of available computing resources e.g., by consolidating certain VMs in a specific host, exploiting better multicore architectures, or interleaving I/O intensive applications with computing intensive ones. In this respect, awareness of energy consumption or environmental impact of computation can encourage users to change their behaviour and consider different factors in application deployment and execution decision making model.

5 CONCLUDING REMARKS

In the paper we have presented the $ECO_2Clouds$ approach to monitoring energy consumption and CO_2 emissions as eco-metrics in a federated cloud environment. We have discussed the potential for exploiting the use of eco-metrics in different directions, ranging from the evaluation of environmental impact of cloud computing to encouraging the changes in behaviour of users.

$ECO_2Clouds$ will also develop sophisticated mechanisms for automatic optimization of application deployment configurations based on context-aware service provisioning in the cloud. Optimization will be carried out both at the scheduler level and at the application level during the execution lifetime of applications.

Experimentation in the project will be aimed at testing different types of application case studies, ranging from computing intensive to more interactive and cloud specific long running applications, to analyse how different factors could have an impact on approaches for assessment of environmental impact and create new opportunities for optimization of energy consumption and CO_2e footprint or for a more informed user behaviour that can lead to further developments and innovation in this area.

ACKNOWLEDGEMENTS

REFERENCES

Beloglazov, A., Buyya, R., Choon Lee, Y., Zomaya, A.Y., 2011. A Taxonomy and Survey of Energy-Efficient Data Centers and Cloud Computing Systems. *Advances in Computers* 82: 47-111

BonFIRE, 2013. http://www.bonfire-project.eu/ link checked on Dec. 29, 2013

Bucchiarone, A., Cappiello, C. Di Nitto, E., Barbara Pernici, Sandonini, A., 2012: A Variable Context Model for Adaptable Service-Based Applications. *IJARAS* 3(3): 35-53

Cappiello, C., Datre, S., Fugini, M. G., Gienger, M., Melia', P., Pernici, B. 2013. Monitoring and Assessing Energy Consumption and CO$_2$ Emissions in Cloud-based Systems, *IEEE International Conference on Systems, Man and Cybernetics (SMC)*

Chen, D., Henis, E., Kat, R. I., Sotnikov, D., Cappiello, C., Ferreira, A. M., Pernici, B., Vitali, M., Jiang, T., Liu, J., Kipp, A., 2011. Usage centric green performance indicators, *Proceedings of GreenMetrics 2011* (SIGMETRICS workshop).

ECO$_2$Clouds, 2013. D3.1: Layered Sets of Metrics and Application Profile, ECO$_2$Clouds project deliverable, http://eco2clouds.eu

Greenpeace, 2012. How clean is your cloud?, http://www.greenpeace.org/international/en/publications/Campaign-reports/Climate-Reports/How-Clean-is-Your-Cloud/ link checked on Dec. 29, 2013

Khosravi, A. Garg, S. K., Buyya, R., 2013. Energy and Carbon-Efficient Placement of Virtual Machines in Distributed Cloud Data Centers. *Euro-Par 2013,* 317-328

Kolodziej, J., Khan, S. U., Wang, L., Zomaya, A. Y., 2012. Energy efficient genetic-based schedulers in computational Grids. *Concurrency and Computation: Practice and Experience*

Kipp, A., Jiang, T., Fugini, M. G., Salomie, I., 2012. Layered Green Performance Indicators, *Future Generation Comp. Syst.* 28(2), 478-489

Lindberg. P. et al., 2012. Comparison and analysis of eight scheduling heuristics for the optimization of energy consumption and makespan in large-scale distributed systems. *In Journal of Supercomputing*, Vol. 59, Issue 1, 323-360.

Papazoglou, M. P., Pohl, K., Parkin, M., Metzger, A. (Eds.), 2010. *Service Research Challenges and Solutions for the Future Internet - S-Cube - Towards Engineering, Managing and Adapting Service-Based Systems.* Springer, Lecture Notes in Computer Science

S-Cube Team, http://www.s-cube-network.eu/, link checked on Dec. 29, 2013

Reiter, M., Fettke, P., Loos, P., 2013. Towards a Reference Model for Ecological IT Service Management, *International Conference on Information Systems*, Milan

Singh, H. Reuters, T. (eds), 2011. Data Center Maturity Model, http://www.thegreengrid.org/en/Global/Content/white-papers/DataCenterMaturityModel Dec. 29, 2013

Tenschert, A., Gienger, M., 2013. Cloud Federation Monitoring for an Improved Eco-Efficiency, *eChallenges 2013*, Dublin, Ireland

Wajid, U., Marín, C. A, Mehandjiev, N., 2013. Optimizing Service Ecosystems in the Cloud, in *The Future Internet*, Lecture Notes in Computer Science, Volume 7858, Springer.