The Green Sustainable Telco Cloud

Minimizing greenhouse gas emissions of server load migrations between distributed data centres

Thomas Dandres, Rejean Samson
Polytechnique Montreal, CIRAIG
Montreal, Canada
thomas.dandres@polymtl.ca

Reza Farrahi Moghaddam
Ecole de technologie superieure, Synchromedia
Ericsson Canada Inc.
Montreal, Canada

Kim Khoa Nguyen, Mohamed Cheriet
Ecole de technologie superieure, Synchromedia
Montreal, Canada

Yves Lemieux
Ericsson Canada Inc.
Montreal, Canada

Abstract*— Among the innovative approaches to reduce the greenhouse gas (GHG) emissions of data centres during their use phase, electrical power from renewable sources appears promising. However, renewable electricity is often intermittent due to meteorological conditions. Consequently, the regional availability of renewable power varies constantly over time. This created the opportunity to deploy cloud computing systems relying on data centres located in different regions. Cloud computing technology enables real-time load migration to a data centre in the region where the GHG emissions per kWh are the lowest. While this approach is becoming popular to manage distributed data centres, there is still room for improvement in its implementation. Indeed, the consequences of data centre power demand migrations across electric networks and the resulting GHG emissions are usually neglected. In this project, we developed a novel GHG emission factor based on the sources of electricity affected by the server load migrations. Then, we used this emission factor in a simulation of distributed data centres to minimize their GHG emissions. Results show, the use of the novel emission factor enables an extra reduction of 23% of GHG emissions as compared to the usual approach.

***Keywords*—Data centre network optimization, real-time electricity generation, Green house gas emissions**

I. INTRODUCTION

Information and communication technologies (ICTs) have grown exponentially in the last decades and this rapid growth is expected to continue [1]. However, ICT solutions are associated with the consumption of large amounts of electricity during the use phase [2]. In 2006, ICTs were found to contribute to 2% of global anthropogenic greenhouse gas (GHG) emissions, which was equivalent to the emissions of the aviation industry [3]. Because data centres are one of the three major sinks of electricity among ICT infrastructures, they also significantly contribute to ICT GHG emissions [2]. Therefore, significant effort has been invested to curb data centre electricity demand, improve data centre efficiency and reduce the data centres' environmental footprint [4-15].

Among the innovative approaches to reduce the GHG emissions of data centres in their use phase is overall load management across distributed data centres [4, 6-8, 10, 11, 13-20]. In this approach, data centres are located in several regions and connected to the regional electrical grid. Load management is used to vary the power demand of the data centres in real time in order to maximize power consumption in regions where the GHG emissions per kWh are the lowest at any given time. Indeed, electricity is generated to instantly meet the regional power demand, which changes continuously during the day depending on consumer needs. Therefore, the regional mix of power plants changes constantly and so does the related GHG emissions factor per kWh.

Based on our knowledge, in the conventional approach, the choice of region to which the load is migrated has always been in relation to real-time electricity generation data (in the best cases). Concretely, the electrical grid mixes of each region are checked regularly and then the load is balanced between regions where the GHG emissions are lower at a given time. An issue is that the changes in regional power demand caused by the load balancing are not taken into account in the conventional approach since the migrations are made after the grid mix check. Consequently, the change in each regional GHG emission factor directly caused by load management is also ignored. In other words, the conventional approach would not capture an increase in regional coal power generation (and related emissions) caused by a rise in the load processed by a data centre connected to the electrical grid. Thus, there is uncertainty regarding the real GHG emissions reductions achieved (if any) with the optimization of data centre networks when using the conventional load balancing approach. Therefore, a method adapted to the dynamic electricity context is needed to instantly minimize the GHG emissions of server load migrations within a data centre network connected to the electrical grid.

Thus, the main objective of this study is to develop a new emission factor taking into account the power generation technologies affected by load migrations between data centres of different regions in order to contribute to minimize distributed data centres GHG emissions. The second objective is to evaluate the GHG emissions that can be reduced by server load migrations when compared to a non-cloud situation or to a conventional management of server load migrations. For the

978-3-901882-85-2 (c) 2016 IFIP

purpose of illustration, this research refers to a case study in which the GHG emissions of a cloud computing service are minimized using load balancing between several data centres located in different Canadian provinces.

II. Method

A. The Green Sustainable Telco Cloud

The Green Sustainable Telco Cloud (GSTC) aims to be a cloud computing service based on an efficient, optimized and environmentally-friendly data centre network. Several optimization criteria such as service quality and operating costs are considered but this paper focuses only on the environmental criteria in order to fully illustrate the use of an emission factor involving the power generation technologies affected by the load migrations to mitigate the GSTC GHG emissions.

B. Description of the case study

The purpose of the case study is to illustrate the use of an emission factor capturing the power generation technologies affected by the load migrations when computing the carbon footprint of a service provided by distributed data centres. Thus, it is not presenting accurate and complete carbon footprint results. For this reason, the case study was deliberately simplified and is more conceptual than practical. In this case study, two virtualized data centres are located in the Canadian provinces of Ontario and Alberta and form a cloud computing system that provides online service to populations of these provinces. Ontario and Alberta were chosen because detailed, historic and real-time electricity generation data are available from public utility websites [21, 22].

To simplify the calculations, it is assumed that the two data centres are similar and connected to the regional grid and that only one handles the cloud computing service at a time. In addition, the power demand related to the cloud varies over time on a daily basis depending on user requests, as presented in Figure 1 Since the two data centres are similar, their electricity consumption to process the cloud computing service is the same and the sum of the power demand required by the two data centres is always equal to the cloud power demand in Figure 1. It was also assumed that there are two million users located in Ontario and Alberta and that data transmission by all users towards the data centre in the cloud requires the same electricity consumption, regardless of the location of the users.

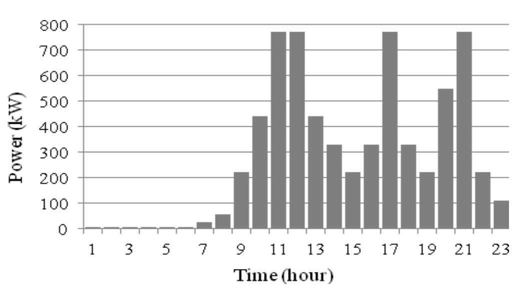

Fig. 1. Hourly cloud power requirement

It was considered that one load migration between the two data centres could be made every hour. Assuming that the GSTC provides an online service that host a negligible amount of user data, the load migrations should not cause significant additional data traffic, and the electricity consumption due to such data traffic can be neglected. Concretely, the GSTC service could consist of online file processing such as picture or video editing or mathematical computing.

C. Identification of affected sources of electricity

The data collected from the Canadian utilities quantify the amount of electricity generated per power plant and per hour in Alberta and Ontario in 2012. Since the emission profile of each power plant is not available, all the data corresponding to the same electricity generation technology were aggregated. Then, the variation in generation capacity of each technology was monitored for each hour in Alberta and Ontario in 2012. The result is an increase or a decrease in power generation per technology, per hour and per Canadian province in response to local power demand change. It constitutes the hourly power generation technologies affected by power demand changes.

One simplification in the identification of the affected technologies is that electricity imports were not considered. Thus, in these simulations, a local change in the power demand can only affect local power plants. However, it is not always true in reality since it may be cheaper to import electricity than to operate locally an expensive power plant to supply an extra local power demand. Alberta and Ontario electricity imports being small compared to their domestic power generation, this assumption is pretty realistic.

D. Calculating the GHG emissions factor

Once the affected power generation technologies are identified for every hour of 2012, the GHG emissions are computed for this granularity based on the amount of electricity provided at each hour by each type of affected power generation technology and on the life cycle GHG emission factor related to each technology. The life cycle emissions factors were taken from the ecoinvent database (version 3.1) using Simapro software (version 8.1) for 1 kWh of electricity for each power generation technology based on the IMPACT2002+ (2.20) impact assessment method. Life cycle emissions factors account for power plant construction as well as other life cycle steps in electricity generation and are not restricted to energy extraction from fossil/fissile fuels or the transformation of renewable energy into electricity. Hourly GHG emissions attributed to affected power generation technologies were then divided by the amount of electricity generated the affected technologies during the hour to obtain the hourly GHG emissions factor per kWh.

E. Real-time GHG emissions of server load migrations

In the context of the case study, the GHG emissions of the server load are computed for every hour of 2012. The total GHG emissions (E) are obtained with equation 1:

$$E = \sum_{h=1}^{8760} F_h \times C_h \qquad (1)$$

Where:

Identify applicable sponsor/s here. If no sponsors, delete this text box (*sponsors*).

F_h is the emission factor of the region where the server load is processed at hour h;

C_h is the electricity consumed by the cloud at hour h.

The region where the load is processed at a given time is the region where affected power generation technologies emit the least amount of GHGs per kWh at this time (i.e. the region with the lowest GHG emission factor). It should be noted that the choice of the region where the server load is migrated is made with a perfect knowledge of the emission factors. Indeed, the emission factors are computed for a past year. Thus this approach makes it possible to evaluate the potential of GHG emissions that can be minimized by managing server load migrations and to compare it with alternative situations.

F. *Comparison with a no-cloud computing scenario*

To assess the benefits of the load migrations between the two Canadian data centres, two scenarios without cloud computing systems were modelled. In these scenarios, it is presumed that the online service is hosted by only one data centre. The data centre is located in Ontario in one scenario and in Alberta in the other. Data centre GHG emissions in 2012 were calculated for each scenario based on the emission factors attributed to affected power generation technologies computed previously. Then, the GHG emissions of the three scenarios (Ontario and Alberta, only Ontario and only Alberta) were compared.

G. *Comparison with the conventional approach*

A comparison with the conventional approach highlights the differences in GHG emissions accounting when considering (or not) the power generation technologies affected by load migration. To this end, the GHG emissions are calculated using average GHG emission factors in equation 1. These average GHG emission factors were computed for Ontario and Alberta for every hour of 2012 based on the global hourly electricity mix (instead of the hourly mix of affected power generation technologies). Then, these regional average GHG emission factors were compared for every hour of 2012 in Ontario and Alberta. Following the principle of the conventional approach, the server load is always processed in the region where the average GHG factor is the lowest. It was also considered that one load migration between the two data centres could be made every hour . Finally, the GHG emissions of the data centre in the two scenarios without cloud computing systems were computed using the hourly regional average GHG emissions factors to have a common basis for discussion with the proposed approach.

III. RESULTS

A. *Real-time GHG emissions of server load migrations*

GHG emissions per kWh and per context (cloud computing and non-cloud computing scenarios) are presented in Table I. As illustrated, cloud computing helps minimize the online service GHG emissions. This reduction in GHG emissions is possible because GHG emissions maximums do not occur at the same time in Alberta and Ontario (especially since the provinces are in different time zones). By avoiding the emission maximums of both regions, the mean of GHG emissions per kWh decreases in the cloud computing scenario as compared to no cloud computing. Concretely, the consideration of power generation technologies affected by server load migrations makes it possible to reduce GHG emissions by 23% as compared to the scenario in which a single data centre is located in Ontario and by 44% when a single data centre is located in Alberta. These results corroborate that there is a high potential of GHG emissions to be avoided when load migrations are managed properly, even when power generation technologies affected by load migrations are considered.

The difference in GHG emissions reduction between Ontario and Alberta is explained by the difference in technologies affected by power demand changes in these regions. While affected power generation technologies rely usually on natural gas and hydro (low GHG emissions per kWh) in Ontario, affected technologies in Alberta generally relies on natural gas and coal (high GHG emissions per kWh). Therefore, the mean GHG emissions per kWh in Alberta are higher than in Ontario. The GSTC is hosted more often in Ontario (71% of the time) than in Alberta. These results highlight the need to check the regional power generation technologies affected by power demand changes before deploying a cloud computing system to minimize GHG emissions (when the cloud is supported by data centres located in different regions).

TABLE I. GHG EMISSIONS PER KWH AND PER SCENARIO

GHG emissions (kg CO2 eq.)	**Server load migrations**	**Ontario**	**Alberta**
	(cloud computing)	***(no cloud computing)***	
per kWh:			
Mean	0.277	0.343	0.543
Standard deviation	0.195	0.218	0.289
per scenario:			
Total	676,696	875,581	1,206,783

B. *Comparison with the conventional approach*

In the conventional approach, the regional average electricity grid mixes were compared in real time to migrate the load to a data centre in the region that has the lowest GHG emissions per average kWh (as opposed to kWh generated by effected power generation technologies). GHG emissions per average kWh and per situation (cloud computing and non-cloud computing scenarios) are presented in Table II. According to the conventional approach, the minimization of the GHG emissions of the cloud computing service involves processing the load in the data centre located in Ontario at all times (*Optimization* and *Ontario* columns are identical in Table II). This result was expected since nuclear power (low GHG emissions per kWh) is the main source of electricity in Ontario versus coal power in Alberta (high GHG emissions per kWh). The maximum GHG emissions per average kWh in Ontario are therefore lower than the minimum GHG emissions per average kWh in Alberta and the data centre in Alberta is never chosen to process the load. As a result, according to the conventional approach, the data centre located in Alberta is irrelevant. This leads to an important finding: the conventional approach does not always yield a fully optimized solution (regardless of the fact that the conventional approach may not accurately model the GHG emissions).

TABLE II. GHG EMISSIONS PER AVERAGE KWH AND PER SCENARIO

Average GHG emissions (kg CO2 eq.)	Optimization	Ontario	Alberta
	(cloud computing)	*(no cloud computing)*	
per average kWh:			
Minimum	0.044	0.044	0.790
Mean	0.150	0.150	0.947
Maximum	0.329	0.329	1.049
Standard deviation	0.058	0.058	0.035
per scenario:			
Total	419,222	419,222	2,303,829

Regarding the emissions presented in Table I and II, GHG emissions per kWh generated by affected power generation technologies in Ontario are higher as compared to GHG emissions per average kWh in Ontario. The main reason is that nuclear power contributes very little to the affected power generation technology mix but represents about half of the power generation of the average power mix. Since nuclear power emits very few GHG emissions per kWh, the small contribution of nuclear power to the affected power generation technology mix makes the GHG emissions per kWh generated by affected power generation technologies greater than the GHG emissions per average kWh. In Alberta, the GHG emissions per kWh generated by affected power generation technologies are lower than the GHG emissions per average kWh because many Alberta power plants burn coal continuously without considering the province’s power demand. These power plants are therefore excluded from Alberta’s affected power generation technology mix. Since coal power plants release significant amounts of GHG per kWh, the affected power generation technology mix in Alberta emits fewer GHG emissions per kWh than the average power mix.

IV. DISCUSSION

The consideration of the power generation technologies affected by load migrations in the computing GHG emissions in real-time is expected to better represent reality and thus to improve the minimization of distributed data centres GHG emissions. However, the implementation of an emission factor considering load migration externalities is not obvious in the context of the current GHG emissions accounting framework. Indeed, currents standards and guidelines issued by the GHG Protocol, ETSI or ITU for ICT assessment are based on average electricity generation emission factors and do not consider power generation technologies affected by data centre power demand migrations. In certain cases, as the distributed data centres in Ontario and Alberta, application of the current standards and guidelines results in an under optimized situation and prevents the consideration of the affected power generation technologies. Indeed, if the data centre located in Alberta is chosen to reduce GHG emissions from affected technologies, the standards and guidelines would claim high GHG emissions because of the average power mix of Alberta (which includes mostly technologies not affected by the load migration). Therefore, there is a need to adapt the standards and guidelines in order to confer an official recognition to those who manage affected power generation technologies to minimize their GHG emissions.

In this study, the simulation was made on a past period. Thus, it was easy to retrospectively identify the affected power generation technologies. However, to implement the proposed approach in reality, it is needed to predict the affected power generation technologies in real-time. Such predictions may be pretty uncertain due to the high complexity of electric networks. Indeed, it is observed that the affected power generation technology mix varies greatly over time. To effectively reduce GHG emissions of distributed data centres by managing affected power generation technologies, it is needed to reach a certain level of validity in their prediction.

V. CONCLUSION

This paper highlights the role of power generation technologies affected by server load in the real-time optimization of multiregional data centre networks. It identifies an important issue regarding the conventional approach based on load balancing to minimize GHG emissions: the technologies used to generate electricity supplying the data centres that process the load after load migration are ignored. Therefore, the GHG emissions reductions claimed by the conventional approach (if any) are uncertain. This study shows that the real-time management of affected power generation technologies that power two data centres located in two different regions can lead to significant reductions in GHG emissions as compared to a single data centre solution. Interestingly, the identification of the affected power generation technologies leads to GHG emissions reductions that would not have been possible using the conventional approach. Indeed, regional GHG emissions per kWh are quite different when considering affected or average technology mix.

However, considering technologies affected by server load migrations requires predictions of electric grid behaviour, which are uncertain if the electric utilities do not make the information public in real time. More work in needed to improve prediction models based on historical data when planned power generation data are not available. Moreover, the introduction of affected power generation technologies in the optimization of distributed data centres conflicts with the current GHG emission accounting standards and guidelines. Indeed, depending on the emission assessment approach, the GHG emissions per kWh generated by power generation technologies not affected by load migration in a region may be so great that they would prevent a cloud computing service to rely on a regional data centre (even on a temporary basis) even though electricity generated by affected power generation technologies in the region would emit fewer GHG emissions than other considered regions. The solution of this conflict probably lies in the adaptation of the GHG emission accounting standards and guidelines so they can confer an official recognition when affected power generation technologies are managed to minimize distributed data centres emissions.

Finally, while the cloud computing technology enables load migrations leading to potential reductions in GHG emissions, significant efforts should also be invested to make electricity generation less harmful to the environment.

ACKNOWLEDGMENT

The authors would like to thank NSERC and industrial partners Ericsson-Montreal and Inocybe Technologies for their financial support under grant CRDPJ 424371-11.

REFERENCES

[1] J. Gantz and D. Reinsel, "The Digital Universe in 2020," IDC Analyse the Future2012.

[2] B. Lannoo, S. Lambert, W. Van Heddeghem, M. Pickavet, F. Kuipers, G. Koutitas, et al., "Overview of ICT energy consumption," Network of Excellence in Internet Science2013.

[3] The Climate Group, "SMART 2020: Enabling the low carbon economy in the information age," Global eSustainability Initiative2008.

[4] A. Beloglazov, J. Abawajy, and R. Buyya, "Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing," Future Generation Computer Systems, vol. 28, pp. 755-768, 2012.

[5] D. Chen, E. Henis, R. I. Kat, D. Sotnikov, C. Cappiello, A. M. Ferreira, et al., "Usage centric green performance indicators," SIGMETRICS Perform. Eval. Rev., vol. 39, pp. 92-96, 2011.

[6] C. Chen, B. He, and X. Tang, "Green-Aware Workload Scheduling in Geographically Distributed Data Centers," in IEEE 4th International Conference on Cloud Computing Technology and Science, Taipei, 2012, pp. 82-89.

[7] N. Deng, C. Stewart, D. Gmach, M. Arlitt, and J. Kelley, "Adaptive green hosting," presented at the Proceedings of the 9th international conference on Autonomic computing, San Jose, California, USA, 2012.

[8] J. Doyle, R. Shorten, and D. O'Mahony, "Stratus: Load Balancing the Cloud for Carbon Emissions Control," IEEE Transactions on Cloud Computing, vol. 1, pp. 116-128, 2013.

[9] X. Fan, W.-D. Weber, and L. A. Barroso, "Power provisioning for a warehouse-sized computer," SIGARCH Comput. Archit. News, vol. 35, pp. 13-23, 2007.

[10] P. X. Gao, A. R. Curtis, B. Wong, and S. Keshav, "It's not easy being green," SIGCOMM Comput. Commun. Rev., vol. 42, pp. 211-222, 2012.

[11] M. Ghamkhari and H. Mohsenian-Rad, "Optimal integration of renewable energy resources in data centers with behind-the-meter renewable generator," in Communications (ICC), 2012 IEEE International Conference on, 2012, pp. 3340-3344.

[12] I. Goiri, K. Le, T. D. Nguyen, J. Guitart, J. Torres, and R. Bianchini, "GreenHadoop: leveraging green energy in data-processing frameworks," presented at the Proceedings of the 7th ACM european conference on Computer Systems, Bern, Switzerland, 2012.

[13] W. Van Heddeghem, W. Vereecken, D. Colle, M. Pickavet, and P. Demeester, "Distributed computing for carbon footprint reduction by exploiting low-footprint energy availability," Future Generation Computer Systems, vol. 28, pp. 405-414, 2012.

[14] A. Krioukov, C. Goebel, S. Alspaugh, Y. Chen, D. E. Culler, and R. H. Katz, "Integrating Renewable Energy Using Data Analytics Systems: Challenges and Opportunities," IEEE Data Eng. Bull., vol. 34, pp. 3-11, 2011.

[15] U. Mandal, M. Habib, Z. Shuqiang, B. Mukherjee, and M. Tornatore, "Greening the cloud using renewable-energy-aware service migration," Network, IEEE, vol. 27, pp. 36-43, 2013.

[16] C. Despins, F. Labeau, T. L. Ngoc, R. Labelle, M. Cheriet, C. Thibeault, et al., "Leveraging Green Communications for Carbon Emission Reductions: Techniques, Testbeds, and Emerging Carbon Footprint Standards," IEEE Communications Magazine, pp. 101-109, 2011.

[17] A. Amokrane, M. F. Zhani, R. Langar, R. Boutaba, and G. Pujolle, "Greenhead: Virtual Data Center Embedding across Distributed Infrastructures," Cloud Computing, IEEE Transactions on, vol. 1, pp. 36-49, 2013.

[18] I. Goiri, L. Kien, M. E. Haque, R. Beauchea, T. D. Nguyen, J. Guitart, et al., "GreenSlot: Scheduling energy consumption in green datacenters," in High Performance Computing, Networking, Storage and Analysis (SC), 2011 International Conference for, 2011, pp. 1-11.

[19] X. Wang, J. Huang, Z. Liu, and X. Liu, "Design of Resource and Power Management Framework and Strategies for Datacenters Powered by Intermittent Renewable Energy," Information Technology Journal, vol. 12, pp. 5432-5439, 2013.

[20] F. Farrahi Moghaddam, R. Farrahi Moghaddam, and M. Cheriet, "Carbon-aware distributed cloud: Multi-level grouping genetic algorithm Cluster Computing," Cluster Computing, vol. 18, pp. 477-491, 2015.

[21] AESO. (2013). Market participant information - Data Requests. Available: http://www.aeso.ca/downloads/Hourly_Generation_by_Fuel.pdf

[22] IESO. (2013). Generator Output and Capability. Available: http://reports.ieso.ca/public/GenOutputCapability/