Sustainable Computing: Informatics and Systems 5 (2015) 31–40

Contents lists available at ScienceDirect

Sustainable Computing: Informatics and Systems

journal homepage: www.elsevier.com/locate/suscom

Scheduling workloads in a network of datacentres to reduce electricity cost and carbon footprint

Trung Le[1], David Wright*

Telfer School of Management, University of Ottawa, 55 Laurier East, Ottawa, ON K1N 6N5, Canada

ARTICLE INFO

Article history:
Received 25 February 2013
Received in revised form 12 July 2014
Accepted 18 August 2014

Keywords:
Network of datacenters
Carbon cost
Electricity cost
Job scheduling
Carbon credits and taxes

ABSTRACT

This paper quantifies the extent to which the scheduling of workloads among a network of datacenters can reduce both electricity cost and carbon footprint. Based upon empirical data from California, Alberta and Ontario, it develops an optimization model that quantifies the savings in relation to the price of carbon on carbon markets and in carbon taxes. Combining the electricity cost with the carbon footprint using the price of carbon, results indicate a simultaneous saving of both 8.09% of electricity cost and 11.25% of carbon footprint, when jobs are scheduled in the current time-period. When jobs can be scheduled in future time-periods, a simultaneous saving of both 51.44% of electricity cost and 13.14% of carbon footprint was obtained. These results are shown to be robust with respect to variations in the price of carbon in taxes and markets in the European Emissions Trading System, Australia, British Columbia, California, and Japan, apart from exceptional periods when the carbon price was very low. The paper shows how a cloud operator can demonstrate that these savings are "additional" to business as usual so as to sell carbon credits on a carbon market, and indicates the standards available for certifying and auditing those emissions reductions.

1. Introduction

The rapid growth of cloud computing stimulates the demand for datacentres whose energy cost and associated carbon emissions are therefore also rising. Worldwide, spending on datacentre power and cooling exceeds $30 billion, comparable to spending on new server hardware [1]. Performance per dollar of server cost is increasing more rapidly than performance per watt [2]. Datacentre carbon footprint from purchased electric power was estimated at 80 MtCO2e in 2007 [3], and is forecast to reach 260 MMtCO2e (million metric tonnes of CO_2 equivalent) by 2020 [4].

One approach to controlling electric power use in datacentres is the development of standards for monitoring and controlling energy use. The ITU standard L.1300 [5], gives best practices for mechanical, electrical and computer systems in datacentres, and L.1400, L.1410, L.1420 [6–8] focus on the environmental impact of information and communications technologies in general, from the points of view of technology, product life cycle and applications in non-ICT organizations, respectively. More detailed work specific to datacentres started in the fall of 2012 jointly between ISO and IEC under the general heading "Sustainability for and by IT". It is planned to result in the development of 4 standards: ISO/IEC 30131 [9], 30132 [10], 30133 [11] and 30134 [12], and includes not only the widely used PUE[2] measure, but also ITEE,[3] ITEU[4] AND REF.[5] A major input to this international standardization effort is the Canadian standard [13], which was developed as a result of the Greenstar Project [14].

Energy cost and carbon emissions can be further reduced by combining energy efficiency measures with workload scheduling policies that allocate jobs to datacentres that have low dollar and/or carbon costs. The current paper develops such an optimization model and demonstrates it on data from California, Alberta and Ontario in order to quantify empirically the extent of the dollar and environmental benefits that can be obtained from workload scheduling.

The objective function in our model allows weighting of dollar and environmental costs so that it can be used to optimize dollar costs alone, carbon costs alone or a combination of the two in which dollar and carbon costs are combined into a single measure using

---

* Corresponding author. Tel.: +1 613 562 5800x4784.
E-mail addresses: dwright@uottawa.ca, wright@management.uottawa.ca (D. Wright).

[1] Present address: Guidewire Software, 1001 E Hillsdale Blvd #800, Foster City, CA 94404, United States.

[2] PUE, Power Usage Efficiency.

[3] ITEE, IT Equipment Efficiency.

[4] ITEU, IT Equipment Usage.

[5] REF, Renewable Energy Factor.

http://dx.doi.org/10.1016/j.suscom.2014.08.012
2210-5379/© 2014 Elsevier Inc. All rights reserved.

a price of carbon. The model can therefore be used by organizations operating in jurisdictions with a carbon tax (such as British Columbia), those with a carbon price set by a trading scheme (such as California, China and the European Union) and by datacentre operators and users whose trade-off between dollar and carbon costs is subjective.

The model also allows scheduling in both time and space, so that a non-interactive workload can be run up to 24 h in the future, in addition to a choice of the location of the datacentre. Electricity dollar price varies with time of day and with geographical location. The minimum and maximum prices can differ by a factor of 2 and prices in different regional markets are not well correlated, even when nearby [15]. Carbon cost depends on the generating mix among coal, nuclear, natural gas, hydro, wind and solar, which again is very different in different locations and times of day.

We first present a review of previous work in this area, and indicate the way in which this paper extends it. We then describe our optimization model and show how it is applied to empirical data from Ontario, Alberta and California to reduce electricity costs and/or carbon emissions. The data sources for the model are then given and the results presented and discussed in the context of carbon pricing.

2. Previous work

Qureshi et al. [15] provide a method for mapping client requests to geographically distributed datacenters so that the overall electricity dollar cost is minimized. They apply it to a request trace from a large content distribution network using hourly electricity prices and find that electricity costs can be reduced by up to 40%. They suggest that future work in this area includes minimizing environmental cost in addition to dollar cost, an extension provided by the present paper.

Le et al. [16] extend the work of Qureshi by minimizing electricity dollar costs subject to caps on brown (i.e. carbon-intensive) energy use, thus representing the operation of datacenters in jurisdictions where government regulators may impose caps on their carbon emissions. Simulating a search engine trace on datacenters in Europe and the US West and East Coasts, they find that it is possible to reduce brown energy usage by 24% at a dollar cost increase of only 10%. Their approach differs from ours in that they regard carbon emissions as a constraint, whereas we regard it as part of the objective function. To date government regulators have not imposed caps on indirect carbon emissions from datacenters, concentrating instead on direct emissions from heavy industrial emitters such as cement plants. For instance, Korea plans to impose caps on 450 facilities with direct emissions over 25,000 tonnes of carbon per year. The advantage of regarding carbon emissions as an objective to be minimized is that it enables the cloud computing operator to sell carbon credits or offsets by reducing emissions to an economically efficient level determined by the current price of carbon.

Rao et al. [17] optimize electricity cost in a network of datacenters in regions where the wholesale price varies on an hourly basis in some markets and remains constant in others. They use a constrained mixed-integer programming formulation, which is computationally intensive to solve in real time and therefore approximate the solution using linear programming. They find that their proposed algorithm can reduce the total electricity cost up to 30.15% in a simulation involving two locations with time varying prices and one with a fixed price.

Garg et al. [18], schedule high performance computing workloads among datacenters using a multi-objective approach, first choosing datacenter(s) that minimize carbon emissions and then minimizing electricity dollar cost among that group of datacenters. They achieve a 33% dollar cost reduction using annual averages of electricity prices. A more dynamic approach using time-varying wholesale electricity prices and scheduling workloads in time as well as geographic location could potentially achieve further reductions, and is pursued in the current paper.

Narayan and Sharangi [19] introduce a factor not considered by previous authors, namely the heterogeneity of datacenters. This is important with interactive workloads since the latency varies across datacenters, and their best scheduling algorithm combines electricity dollar cost and latency to achieve a 21% energy saving. The focus on latency implies that it is not possible to delay a workload until a time when electricity price is lower. The authors identify including carbon emissions as a possible future extension of their work.

Liu et al. [20] adopt a novel approach to combining environmental and economic objectives by dynamically pricing electricity in proportion to the fraction of energy coming from non-renewable sources. This results in 40% cost saving during times of light traffic.

In the current paper, datacenters obtain their power from the grid and the carbon intensity of that power is obtained from empirical sources. Lin et al. [38] and Van Heddeghem et al. [39] address the situation with solar and wind power generation located locally at the datacenters. Lin et al. [38] set the capital and marginal cost of renewables to zero and focus on minimizing a combination of the electric power cost and the cost of delaying the execution of delay-sensitive jobs. Van Heddeghem et al. [39] focus on migrating jobs among datacenters according to the local availability of wind and solar power similar to the Greenstar Project [14] and hence emphasize the importance of having spare capacity in terms of servers and datacenters. The carbon cost of manufacturing this spare capacity is therefore also included in their calculations and is calculated to be about 9% of the total carbon cost of building plus operating a datacenter.

The work described above is similar to the work described in the current paper in that it deals with the scheduling of workloads among a geographically distributed set of datacenters. Additional savings can be achieved by optimization within a single datacenter. Since this is not the focus of the current paper, we refer briefly to only a few recent papers. Liu et al. [37] optimize electric power cost and measure the associated carbon emissions but do not include carbon in the objective function. Moghaddam et al. [40] model the energy consumption and carbon footprint of a single virtual machine. Chen et al. [41] reduce server energy consumption by shutting down selected servers whilst maintaining performance-based service level agreements. Mazzucco and Dyachuk [21], schedule the number of servers in a datacenter to be powered down when serving a workload that varies over time. These studies reduce the electric power consumption of a single datacenter, so that carbon emissions are automatically also reduced by the same percentage according to the generating mix of the power supply for that datacenter. Carbon emissions cannot therefore be treated as a separate objective or constraint.

A summary comparison indicating the different approach of each author and the contribution of the current paper is provided in Table 1.

3. Optimization model

Fig. 1 illustrates a scheduler in a cloud computing service that distributes jobs among datacenters at N geographical locations over K time periods. The datacenters report to the scheduler: their spare capacity (in CPU hours), their electricity prices (in $ per kWh) and their emissions intensity (in tonnes of CO_2e per kWh) over the next K time periods. The cloud management may specify the weights to be attached to electricity price and emissions intensity, for instance determined by the current price of carbon, or according to their subjective preferences.