

Cost-Aware Building of Cloud Services Powered by Renewable Energy

Josep Ll. Berral[†], Íñigo Goiri, Thu D. Nguyen, Ricard Gavaldà[†], Jordi Torres^{†‡}, Ricardo Bianchini
{berral,torres}@ac.upc.edu, gavald@lsi.upc.edu, {goiri,tdnguyen,ricardob}@cs.rutgers.edu

[†]Universitat Politecnica de Catalunya, [‡]Barcelona Supercomputing Center, Rutgers University

Technical Report DCS-TR-706, Dept. of Computer Science, Rutgers University, December 2013

Abstract—Interest in powering datacenters at least partially using on-site renewable sources, e.g. solar or wind, has been growing. In fact, researchers have studied distributed services comprising networks of such “green” datacenters, and load distribution approaches that “follow the renewables” to maximize their use. Unfortunately, the prior works have not considered where to site such a network for efficient production of renewable energy, while minimizing both datacenter and renewable plant building costs. Moreover, researchers have not built real load management systems for follow-the-renewables services. Thus, in this paper, we propose a framework, optimization problem, and solution approach for siting and provisioning green datacenters for a follow-the-renewables HPC cloud service. We illustrate the location selection tradeoffs by quantifying the minimum cost of achieving different amounts of renewable energy. Finally, we design and implement a system capable of migrating virtual machines across the green datacenters to follow the renewables. Among other interesting results, we demonstrate that one can build green HPC cloud services at a relatively low additional cost compared to existing services.

I. INTRODUCTION

Motivation. It is well known that datacenters consume an enormous amount of electricity. Estimates for 2010 indicate that datacenters consume around 1.5% of the total electricity used world-wide [18]. This consumption translates into high carbon emissions, since most of this electricity is produced using fossil fuels. A 2008 study estimated world-wide datacenters to emit 116 million metric tons of carbon, slightly more than the entire country of Nigeria [25].

With increasing societal awareness of these emissions, there is increasing demand for cleaner products and services. As a result, several companies have built or announced plans to build “green” datacenters, i.e. datacenters with on-site green power plants. For example, Apple [4] and McGraw-Hill [26] have built 20MW and 14MW solar arrays, respectively, for their datacenters. A few small cloud service providers, e.g. Green House Data [15], AISO [1], and GreenCloud [16], operate datacenters powered completely by on-site solar and/or wind farms. As carbon taxes and cap-and-trade start to migrate from Europe and Asia to the US, even more companies are likely to build green datacenters. For example, the UK government created a mandatory cap-and-trade scheme for businesses consuming more than 6 GWh per year [40]; i.e., a company with even a small 700-kW datacenter must participate.

Some research efforts have explored how to incorporate

sources of renewable (“green”) energy such as solar and wind into datacenters. In particular, several studies have considered load distribution between geographically distributed datacenters to take advantage of green energy produced on-site [21], [23], [24], [45]. Two key observations behind these works are: (1) services are often replicated on multiple geographically distributed datacenters for high availability and low response time, and (2) diverse generation of green energy because of differing local weather conditions. Thus, load distribution approaches that consider green energy production can allow the workload to “follow the renewables” to increase green energy usage. However, to the best of our knowledge, no previous work has considered where to site such a network of datacenters globally for efficient production of green energy, while minimizing both datacenter and renewable plant building costs. Moreover, prior works have not built real load management systems for follow-the-renewables cloud services. These are exactly the topics of this paper.

Selecting sites. A service provider seeking to create a network of green datacenters must consider the significant cost of building, provisioning, and operating these datacenters. The capital costs include land acquisition, datacenter and green power plant construction, and bringing enough network bandwidth and electricity to the datacenter. The operational cost includes electricity if the datacenter is not completely powered by the on-site green power plant, networking, and water for cooling.

An interesting aspect of this problem is that the cost of the service can depend heavily on the specific locations of the datacenters. For example, the production of green energy from sources such as solar and wind depends strongly on location. Land and electricity prices, as well as the provisioning of the cooling system, also depend on location. Further, the location-dependent costs are not always directly related. For example, many good locations for solar energy production are hot and so their cooling would be expensive.

Given the pervasive impact of the datacenters’ locations, cloud service providers must intelligently select them. As there can be many potential locations for each datacenter and many issues to consider in evaluating them, the site selection process can be very difficult. Thus, in this paper, we study the site selection and provisioning process, while characterizing a large set of locations across the globe as potential locations