Rendering Green or Clean Energy by Load Balancing in Data Centres

Vikas Kumar
Assistant Professor, Department of Electrical Engineering, Chandigarh University Mohali Punja

Kamal Kant Sharma
Assistant Professor, Department of Electrical Engineering, Chandigarh University Mohali Punjab

Akhil Gupta,
Assistant Professor, Department of Electrical Engineering, Chandigarh University Mohali Punjab

Gurinder Kaur Sodhi
Assistant Professor, Department of Electronics and Communication Engineering, Desh Bhagat University

ABSTRACT
There have been various academic and commercial endeavours for reducing power costs and carbon footprints by applying computing technologies, and more efficient power delivery and cooling systems. However, energy efficiency alone will slow down the growth of IT carbon footprint. To maintain safe levels of global greenhouse gases, renewable energy sources is becoming a prioritized choice for IT companies to power their rapidly expanding data centre infrastructures. The international environmental organization Greenpeace states that "Green IT = Energy Efficiency + Renewable Energy". Thus, many governments enact renewable portfolio standards and provide incentives for green power generation and usage. Additionally, the improvements in power generating efficiency and cost/Watt reductions of renewable energy will reduce the deployment cost significantly in the future. For example, the efficiency of solar panels is expected to triple but the cost/Watt of solar panels is expected to halve in 2030.

Keywords: Green IT, Green Peace, data centre, Renewable Energy, Green ware system.

1. INTRODUCTION
The renewable energy sources, such as solar and wind, to power cloud computing data centers is a hot research topic today. In this paper, summarization of a series of fundamental relevant research problems is done: why, when, where and how to leverage renewable (green or clean) energy in data centers. The development of cloud computing services has promoted massive-scale, geographically distributed datacenters with millions of servers. Large cloud service providers consume many megawatts of power to operate such datacenters and the corresponding annual electricity bills are in the order of tens of millions of dollars — such as Google with over 1,120 GWh and $67 M, and Microsoft with over 600 GWh and $36 M [1]. Reportedly, datacenters now consume about 1.3% of the worldwide electricity and this fraction will grow to 8% by 2020 [2]. As a matter of fact, the aggregate datacenters in the world consumes more electricity than most nations in the world except 4 countries [3]. High energy consumption not only results in large electricity cost, but also incurs high carbon emission. In the United States, generating 1 kWh of electricity emits about 500g of CO_2 on average [2]. Each 100 MW power station will cost $60-100 million dollars to build and emit 50 million tons of CO2 during its operation [3]. As a result, IT carbon footprints currently occupy 2% of global greenhouse gas emissions [4].

2. How green & Clean are the Datacenters
To measure how clean is a datacenter, the Green Grid organization proposes a new sustainability metric, carbon usage effectiveness (CUE), to measure carbon emissions associated with datacenters. CUE is defined as: CUE = Total CO2 Emissions caused by Total Datacenter Energy/IT Energy Consumption. The units of the CUE metric are kilograms of carbon dioxide equivalent (kgCO2e) per kilowatt-hour (kWh). The comparison of carbon emission rate of the most common energy sources is shown in Table I. Table I shows that the renewable energy sources have much less carbon emission rate than fossil fuels such as coal, gas and oil. Even though, fossil fuels still counts for 2/3 of electricity of the world [2].

TABLE I.
Carbon emission rate of major energy sources [2]

Energy Source	Nuclear	Coal	Oil	Hydro	Wind	Solar
Carbon Rate (gCO2e / kWh)	15	968	890	13.5	22.5	53

Large IT companies have started to build data centres with renewable energy, such as Face book's solar-powered data centre in Oregon and Green House Data's wind-powered data centre in Wyoming. In April 2012, Greenpeace released a report asking, "How clean is your cloud?[4]" It examines the data centres built by the large Internet companies, and ranks them according to how efficient their cloud facilities are, and where they get their electricity. It found that both Google (39.4% clean energy) and Yahoo (56.4% clean energy) are active in supporting policies to drive renewable energy investment and to power clouds with green energy. Even worse, there are numerous small and medium-sized data centres that consume the majority of energy, yet much less energy efficient. Apart from the pressures for reducing the huge energy consumption and carbon emission, there are unique opportunities raised by cloud data centres to apply green energy: (1) Cloud service providers typically own geographically distributed data centres. They can distribute workloads among geo-dispersed data centres to benefit from the location diversity of different types of available renewable energies. (2) Cloud data centres usually support a wide range of IT workloads, including both delay-sensitive non-flexible applications such as web browsing and delay-tolerant flexible applications such as scientific computational jobs. The workload flexibility can tackle the challenges in integrating intermittent renewable energy by delaying flexible workloads to periods when renewable sources are abundant without exceeding their execution deadlines. (3) Cloud data centres connect to grids at different locations with time-varying electricity prices. Renewable energy is a means to mitigate the risk against the future rise in power prices. (4) Data centres usually equip with uninterrupted power supply (UPS) in case of power outages. Since UPS battery is usually over-provisioned, UPS can store energy during periods of high renewable generation and supply power when the renewable energy is insufficient.

TABLE II.
Comparison of cost & prices of Energy Sources [3]

Energy Source	Grid	PPA	REC	DG
Cost (P.f/Kwh)	5	6	0.5	30
CER	586	0	0	1056

3. Nerve tract to a Cleaner Cloud

The most common way to utilize renewable energy is to deploy on-site generation equipments at the datacenter facility. Such on-site generation has negligible transmission and distribution loss. However, the location of a datacenter does not necessarily have a profitable on-site renewable energy deployment. Another way is to deploy off-site renewable energy generation plant at the locations with good wind speed or solar irradiation. The generated green energy can be dispatched across the grid to the consuming datacenters [3]. Apart from these explicit options of provisioning renewable energy generation plants, there also exist three kinds of implicit options to utilize renewable energy. (1) Power purchase agreement (PPA), which purchases a portion of green energy from a renewable energy source. (2) Renewable energy credits (RECs), which are tradable, non-tangible energy commodities. They represent that 1 MWh of electricity be generated from an eligible resource. (3) Carbon offsetting, this represents the reduction of one ton of carbon dioxide. Table II shows the costs and carbon emission rate of REC, PPA conventional grid, and diesel generator (DG) used as backup power source in datacenters.

4. Exploring Challenges

The main challenge of utilizing renewable energy is that renewable sources are variable, intermittent and unpredictable. For instance, wind or solar energy is only available when wind blows or sun shines. Characteristics of cloud datacenters further raise challenges to apply renewable energy: (1) Global users require 7×24 cloud services. However, intermittent renewable energy presents a problem for consistent users of power like datacenters. (2) Cloud demand is dynamic, which requires dynamic power provisioning. Thus, the power supply should be elastic. Nevertheless, unlike conventional energy from the grid, renewable energy cannot be scheduled on demand. (3) High reliability services. This incurs the problem of how to construct reliable power supply in the presence of uncertain renewable energy. (4) Automatic management. This requires the power supply system should choose and supply power automatically among multiple power sources. Before deploying renewable energy equipment, we need to know what kinds of datacenter workloads are amenable for green energy and whether the datacenter is suitable for green energy. Afterwards, we need to know the capacity of renewable energy generation for the specific datacenter demand. To mitigate the variability of renewable energy, datacenters can store the generated renewable energy in batteries or on the grid. However, it remains significant challenges when/how to store and supply energy from the battery to meet demand. In addition, these approaches incur energy loss and high additional costs of purchase and maintenance. Instead, we can maximize the use of available energy by matching the demand to the supply. This prompts many interesting research problems. Due to the variable availability of renewable energy, the scheduler has to exploit suitable forecast techniques to predict the available renewable energy in an online fashion.

5. Green energy Generation Model

Wind and solar are the most prominent renewable sources, which currently provide 62% and 13% non-hydro renewable electricity worldwide, respectively [3], can be much lower than grid energy (e.g., 80%). Specifically, the capacity factor of wind energy is within 20%~45%, while the capacity factor of solar energy ranges from 14% to 24% [3]. The solar or wind power output depends on the environmental conditions, such as solar irradiance or wind speed. Consequently, their capacity factor, which is the ratio of the actual output over a period to its potential output if it operated at full nameplate capacity. The wind turbines generation can be modeled as a function of the wind speed. Let v be the wind speed, then the wind power output can be approximated as follows [8]:

$$p_{wind} = \begin{cases} 0 & v < v_{in}, v > v_{out} \\ p_r \cdot \dfrac{v - v_{in}}{v_r - v_{in}} & v_{in} < v < v_r \\ p_r & v_r < v < v_{out} \end{cases} .$$

where vin and vout are the cut-in (typically 3~4 m/s) and cut-out (typically 25 m/s) wind speed, and vr and pr are the nameplate speed and wind turbines power. If a wind farm consists of mw turbine, then the wind power output is the sum of all the turbine power output given by:

$$PW = \sum_{k=1}^{m} P_{wind}$$

where pkwind is the power output of kth wind turbine at the wind speed of v, with the assumption that the wind turbines have the same wind speed in the same wind farm. However, the speed and angle may differ from turbine to turbine in a farm.

6. Planning of Green Datacenter Capacity

Renewable energy can reduce data centre operational cost and carbon footprint when correctly selected. Nevertheless, as illustrated in Fig.1, it is not a trivial task, for the reason that data centre planners should consider different renewable options (on-site, off-site, PPA, RCE), different energy sources (solar, wind), different energy storage devices (batteries, fuel-cells, flywheels), markets prices fluctuations, workload variances, weather conditions, incentives and service penalties. Given all these factors, Jose Renauet.al. [12] Proposed a power simulator that can evaluate the energy cost of a data centre using renewable energy. For any given location and workload of a data centre, one can find the best ratio of renewable energy sources using a genetic algorithm. Differently, a practical linear programming based optimization framework for a data centre to achieve a target carbon footprint at minimal cost, among the various options presented in Fig.1.Under extensive experiments with real-world power demand and renewable energy traces, the key findings are: (1) Renewable energy can lower both carbon emissions and costs for data centres.(2)On-site renewable can lower data centre costs by reducing the peak power draw from the grid. (3) The most cost-effective approaches for carbon reduction vary with different carbon footprint targets. For a moderate target (up to 30%), the best options is using on-site renewable, while a more carbon reduction goal requires off-site renewable, and a zero carbon goal must resort to renewable energy products such as RECs Unfortunately, the model considered a low and constant grid price ($0.05/kWh), which made on-site generation relatively expensive in comparison.

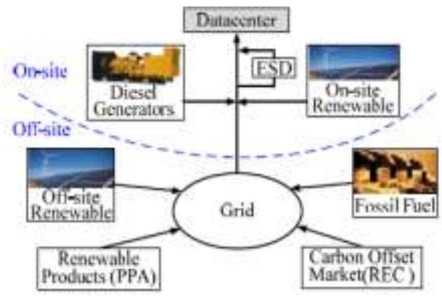

Fig.1 Various renewable energy options for cloud-scale datacenters

7. Inter-Datacenters Load Balancing

Large cloud service providers usually own geo-distributed datacenters for (1) better performance, i.e., close to end users, (2) more reliable system, i.e., multiple backups in different datacenters, (3) lower cost, i.e., leveraging different energy sources, prices, cooling efficiency and tax. The first question is whether geographical load balancing can exploit renewable energy and reduce environmental impact. Free Lunch was proposed to exploit renewable energy [17]. It assumes that datacenters are close to renewable energy sources, and support the seamless execution and migration of virtual machines according to the power availability. High latency and storage synchronization can degrade the efficiency of VM migration. Thus, Free Lunch might be inefficient for latency sensitive or interactive applications. For web requests, Liu et al. exploited whether geographical load balancing can encourage the use of renewable energy and reduce the use of fossil fuel (brown) energy [18]. It was found that if grid energy (with penetration of renewable energy sources) is dynamically priced based on the proportion of brown energy, then geographical load balancing can reduce brown energy usage significantly. Especially, the optimal pricing model at ith datacenter is: $p_i(t) = (1-\alpha_i(t))/\beta$, where $\alpha_i(t)$ the fraction of the energy that is from renewable sources at time t and β is the relative valuation of delay versus energy. Further, based on trace-driven study, they investigated the feasibility of powering datacenters entirely using renewable energy [19]. It observed that "follow the renewable" routing could significantly reduce the required capacity of renewable energy, especially in corporation with suitable size of energy storage. Storage becomes more valuable with higher capacities of renewable, i.e., >1.5, where renewable capacity is defined as the ratio of the average renewable generation to the minimal energy required to serve the average workload. Moreover, they found that wind energy is more viable than solar. The rational is that wind is available during both day and night, and has little correlation across different locations.

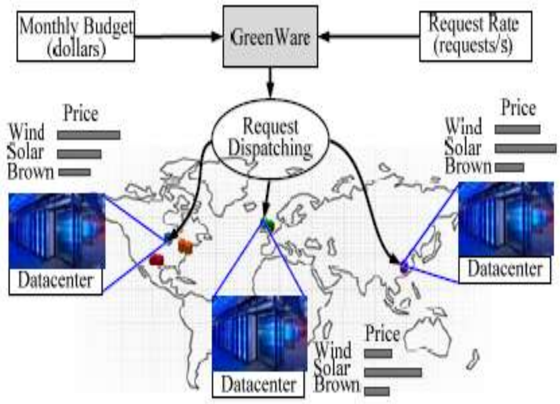

Fig. 2 Green Ware systems for distributed cloud-scale datacenters

The second question is how many and which user requests should be migrated. Stewart et al. [13] proposed an accounting model of per-request energy consumption as a guideline for managing renewable energy. Based on the performance counters, the model collects three metrics including the activities of L2 cache, memory and CPU: L2 cache requests per CPU cycle (Ccache), memory transactions per CPU cycle (Cmem), and the ratio of non-halt CPU cycles (Cnonhalt). The request power model is given as:

Pidle+Pcache.Ccache/Ccellcache+Pmem/Pcell mem+Pnonhalt/Ccell nonhalt

where *P* are the coefficient parameters for the linear model, and *Ccell* are the upper bound for the three metrics. With the model, the power/energy profiles for individual requests can be constructed, and renewable-aware request distribution can be conducted. If incentives for renewable energy usage are not considered, currently renewable energy is often more expensive than brown energy. To utilize renewable energy, cloud service providers have two challenges: (1) how to conduct dynamic request dispatching among geographical datacenters to maximize the use of renewable energy, and (2) how to achieve that within allowed operational cost budget. As illustrated in Fig.2, Green Ware[8] dynamically distributes requests among datacenters to maximize the percentage of used renewable energy subject to the desired cost budget. This is formulated as a constrained optimization problem:

$$\text{Maximize: } \frac{\sum_{i=1}^{N}(W_i + S_i)}{\sum_{i=1}^{N}(W_i + S_i + B_i)}$$

$$\text{s.t. } \sum_{i=1}^{N}(PW_i \cdot W_i + PS_i \cdot S_i + PB_i \cdot B_i) \le Budget$$

where Wi, Si, Bi are the amount of used wind, solar and brown energy in the ith ($1 \le i \le N$) data centre, and PWi, PSi and PBi are the current prices of the three types of energy. Constraint (6) means that the energy cost should be within the cost budget. They proposed an efficient request-dispatching algorithm based on linear-fractional programming (LFP) and implemented Green Ware, based on the linear program solver in Mat lab. Evaluations with real-world weather, electricity price and workload traces show that Green Ware can significantly increase the use of renewable energy without violating the desired cost budget. However, Green Ware ignores a key factor: what set of data users are interested in and where the data is stored. FORTE [2] captures the relationship between user groups, data centres and data. It uses two assignment algorithms to optimally map users to data centres and map data to data centres. The objective is to minimize the weighted sum of access latency, electricity cost and carbon footprint. Trace-driven experiments show that FORTE can reduce carbon emissions by 10% without increasing the electricity bill.

8. CONCLUSION

The key challenge of utilizing renewable energies is the variable, intermittent and unpredictable nature. By presenting a taxonomy of the latest research and exploring new challenges involved in managing the use of uncertain renewable energy, intend to answer why, when, where and how to leverage renewable energy in datacenters. Specifically, we believe that matching uncertain power demand and multi-sources supply in a complementary manner should be one of the highlights in future research.

9. REFERENCES

[1] A.Qurush, “Power-Demand Routing in Massive Geo-Distributed Systems,” Ph.D. dissertation, MIT, 2010.

[2] Peter Xiang Gao, Andrew R. Curtis, Bernard Wong, Srinivasan Keshav, “It's Not Easy Being Green,” Proc. ACM SIGCOMM, Helsinki, Finland, 2012, pp. 211-222.

[3] Chuangang Ren, Di Wang, Bhuvan Urgaonkar, and Anand Sivasubramaniam, “Carbon-Aware Energy Capacity Planning for Datacenters,” Proc. MASCOTS, Washington DC, 2012, pp. 391-400

[4] Cook, G. How clean is Your Cloud? Greenpeace International Tech. Rep., April, 2012

[5] Wiser, R., and G. Barbose, “Renewables Portfolio Standards in the United States — A Status Report with Data Through 2007,” LBNL-154E, 2008.

[6] Bianchini, R, “Leveraging Renewable Energy in Data Centers: Present and Future,” Proc. HPDC’12, San Jose, California, 2012, pp. 135-136.

[7] I. Goiri, R. Beauchea, K. Le, T. Nguyen, M. Haque, J. Guitart, J. Torres, and R. Bianchini, “Greenslot: Scheduling Energy Consumption in Green Datacenters,” Proc. SC’11, Seattle, WA, 2011, pp. 1–11.

[8] Y. Zhang, Y. Wang, and X. Wang, “Greenware: Greening Cloud-scale Data Centers to Maximize the Use of Renewable Energy,” Proc. Middleware, Lisboa, Portugal, 2011, pp. 143–164.

[9] Chao Li, Wangyuan Zhang, Chang-Burm Cho and Tao Li, “SolarCore: Solar Energy Driven Multi-core Architecture Power Management,” Proc. HPCA’11, San Antonio, TX, USA, 2011, pp. 205-216.

[10] B. Aksanli, J. Venkatesh, L. Zhang, and T. Rosing, “Utilizing Green Energy Prediction to Schedule Mixed Batch and Service Jobs in Data Centers,” Proc. HotPower’11, Cascais, Portugal, 2011, pp. 53-57.

[11] I. Goiri, K. Le, T. Nguyen, J. Guitart, J. Torres, and R. Bianchini, “Greenhadoop: Leveraging Green Energy in Data-processing Frameworks” Proc. EuroSys, Bern, Switzerland, 2012, pp. 57-70.

[12] [.M. Brown and J. Renau. Rerack, “Power Simulation for Data Centers with Renewable Energy Generation,” Proc. ACM GreenMetrics, SAN JOSE, CALIFORNIA , 2011, pp. 77-81.

[13] C. Stewart and K. Shen, “Some Joules Are More Precious Than Others: Managing Renewable Energy in the Datacenter,” Proc. HotPower, Montana, USA, 2009, pp. 15-19.

[14] Andrew Krioukov, Sara Alspaugh, Prashanth Mohan, Stephen Dawson-Haggerty, David Culler and Randy Katz, “Design and Evaluation of an Energy Agile Computing Cluster,” University of California, Berkeley, Technical Report, UCB/EECS-2012-13, 2012.

[15] N. Sharma, S. Barker, D. Irwin, and P. Shenoy, “Blink: Managing Server Clusters on Intermittent Power,” ACM SIGARCH Computer Architecture News, vol. 39, no. 1, 2011, pp. 185–198.

[16] Chao Li, Amer Qouneh and Tao Li, “iSwitch: Coordinating and Optimizing Renewable Energy Powered Server Clusters,” Proc. ISCA, Portland, OR, USA, 2012, pp. 512-523.

[17] S. Akash, R. Sohan, A. Rice, A. Moore, and A. Hopper, “Free Lunch: Exploiting Renewable Energy for Computing,” Proc. HotOS, Napa Valley, California, 2011, pp. 17-17.

[18] Z. Liu, M. Lin, A. Wierman, S. Low and L. Andrew, “Greening Geographical Load Balancing,” Proc. SIGMETRICS, San Jose, California, 2011, pp. 233-244.

[19] Z. Liu, M. Lin, A. Wierman, S. H. Low, and L. L. H. Andrew, “Geographical Load Balancing with Renewables,” Proc. ACM GreenMetrics, SAN JOSE, CALIFORNIA , 2011, pp. 62-66.

[20] Wei Deng, Fangming Liu, Hai Jin, Chuan Wu, "SmartDPSS: Cost-Minimizing Multi-source Power Supply for Datacenters with Arbitrary Demand", Proc. IEEE ICDCS, Philadelphia, USA, 2013.