

The Answer Is Blowing in the Wind: Analysis of Powering Internet Data Centers with Wind Energy

Yan Gao
Accenture Technology Labs

Zheng Zeng
Apple Inc.

Xue Liu
McGill University

P. R. Kumar
Texas A&M University

Abstract—Internet-scale data centers (IDCs) have rapidly proliferated to such an extent that their energy consumption and GreenHouse Gas (GHG) emissions have become an important concern to society. As a result, many IDC operators have started using renewable energy, e.g., wind power, to power their data centers. Unfortunately, the utilization of wind energy has stayed at a low ratio due to the intermittent nature of wind. This paper makes the case that it is in fact possible for a distributed IDC system to exploit multiple uncorrelated wind energy sources to significantly reduce the effect of intermittency and nearly achieve “entirely green” cloud-scale services. This result is obtained based on the analysis of real-world wind power traces from 69 wind farms. The idea is to leverage the front-end load dispatching server to send work to the location where wind power is available. We propose a wind-power-aware (WPA) policy that routes jobs based only on the current states of workloads and wind power availabilities in the data centers. We show that with the WPA policy more than 95% of energy consumption in IDCs can in fact be satisfied by wind power, and, secondly, that achieving this does not require the delaying of processing of jobs due to wind availability. We also show that the locations where data centers are placed play an important role in achieving high wind power utilization. Our analysis shows that wind power utilization can generally lie in a range from 44% to 96%, depending on how the locations of wind farms are selected. We propose a method for location selection that uses the coefficient of variation instead of the correlation coefficient, and show that with this method the utilization can lie in the high end of the above range. Finally, we verify these results by simulations that are based on real-world traces for both workloads and wind power generations.

I. INTRODUCTION

With the increased demand for cloud computing services, Internet-scale Data Centers (IDCs) are proliferating worldwide and growing rapidly into massive geographically distributed systems. As a result of the fast growth of the electricity usage in IDCs, concerns about their GHG emissions have risen to both operators and society. To reduce GHG emissions and meet growing energy demands, a potential promising solution is the utilization of renewable on-site energy supply in IDC systems.

Data centers are already starting to add renewable energy into their energy portfolio [1], [2], [3], [4]. Most studies of powering data centers with renewable energy have focused on powering individual data centers with wind or solar power, e.g., [2], [5]. However, the intermittency and unpredictability of renewables make it very challenging to power a data center

entirely by only local wind or solar energy. An emerging solution for this problem is to enable load balancing to be renewable-aware so as to exploit the geographical distributedness of IDCs and increase the renewable energy utilization. This paper investigates this approach concentrating on one type of renewable energy: wind energy.

The goal of this paper is to quantify the benefits of this approach through the analysis of real-world traces of wind power. The questions that this paper answers are: 1) Whether it is possible to achieve “entirely green” IDCs through a wind-aware load balancing design? and 2) How to select data center or wind farm locations to extend the degree to which wind power is consumed by IDCs?

In order to investigate these problems, we consider the following assumptions for IDCs: 1) assume IDCs are a large system composed of multiple fully replicated data centers spread out geographically to wherever wind farms are available; 2) each individual data center performs the ideal workload consolidation and its power consumption is proportional to the workload in the data center; 3) a threshold for the maximum number of jobs that a server can accommodate is set so that QoS is guaranteed if the jobs in a server are less than this threshold; and 4) there is a central workload dispatcher that maps client requests to data centers. We propose a wind-power-aware (WPA) workload mapping policy. It keeps tracks of the ratio of the power consumption to the effective wind power in each data center. Upon the arrival of a new job, the workload dispatcher routes it to the data center that has the minimum ratio. If we consider the ratio as the normalized “queue” length for wind power consumption, then the proposed WPA policy is essentially a variant of the join-the-shortest-queue (JSQ) rule. We build a simulator based on the above model, and emulate the WPA policy to map real-world trace-based workloads to data center while wind power for data centers is also modeled according to real-world traces.

Our study suggests that using this JSQ-based WPA routing can achieve near-to-entirely green Internet-scale services, i.e., more than 95% of energy consumption in IDCs can be satisfied by wind power, and also that attaining this does not require the regulation of workload processing according to the availability of wind power. We find that location selection plays an important role in achieving high wind-power utilization. A bad selection of wind farms can lead to only 44% wind energy utilization. We also discover a relationship between the two statistical coefficients of wind power generation and

This work was supported in part by the National Science Foundation (NSF) under Contracts CNS-1035340 and CNS-1035378.

the wind energy utilization. Surprisingly, our result shows that only depending on correlation coefficients is not sufficient to guarantee a proper wind farm selection. In contrast, we show that the coefficient of variation (CV) of the aggregate wind power is a much better metric because a smaller CV implies less variation, and hence higher availability of wind power. We also quantitatively study the impact of the number of data centers as a function of the given number of servers, and the impact of the number of servers for given workloads. Our analysis suggests that building only a few (≤ 5) data centers can realize most (i.e., 74%-96%) of the improvement in terms of wind power utilization.

In the next section, we provide an overview of the wind power traces. In Section II we present the architecture of the wind-powered IDCs. Section III provides the power consumption model, and in Section IV we describe our wind-power-aware mapping policy. We discuss wind farm selection in Section V. Section VI presents the simulation study.

II. ARCHITECTURE OF WIND-POWERED IDCs

This section introduces the model of the architecture for wind-powered IDCs. The schematic for wind-powered IDCs is illustrated in Figure 1. Servers are spread out over several geographically distributed data centers. Each data center is co-located with an on-site wind farm. The data center draws power directly from the co-located wind farm through a micro-grid system. However, it is not the only energy source. When the local wind power is not enough to match the power demand, the data center can draw power from the electrical grid outside the micro-grid system. For simplicity, we assume that the energy that is drawn from the outside grid is produced with “brown” energy sources such as coal, whereas wind farms provide purely “green” energy.

On top of these distributed data centers, there is a central dispatcher that maps jobs to the data centers. Note that we consider a central dispatcher only for simplicity. The mapping algorithm presented in this paper applies to distributed dispatchers as well. Each client request first inquires with the dispatcher as to which data center it should go to. The dispatcher maps the request to a data center in a real-time manner following the availability of wind power. In this way, the load dynamically shifts from the data center where wind power is unavailable to the one where wind power is available. Note that in this design, data centers do not deliberately hold any jobs expecting the wind power to be available in the future. In fact, not all services are sensitive to the latency and a better approach is to schedule the different jobs according to their QoS requirements. However this is not the focus of this paper and we leave it for future work.

III. MODELING POWER CONSUMPTION

In order to quantify how much power a data center consumes, we need a model to map the workloads to the power consumption. We consider a relatively simple model since our goal is not to provide greatly accurate results, but rather to

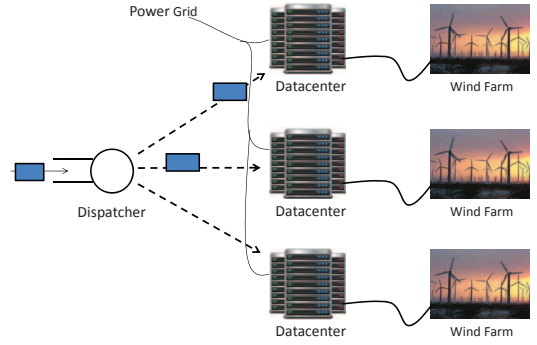


Fig. 1. System Diagram

estimate the wind power utilization. It is worth noting that our analysis can be easily extended to other power models.

We assume that the data center is “power-proportional”, i.e., power use in a data center is proportional to the workloads. A data center can be regarded as a set of servers. However, research [6] has shown that the power usage of a server is not power-proportional, but an affine function of CPU utilization:

$$e = e_0 + e_1 U, \quad (1)$$

where e_0 is the power use when the server is idle, and $e_0 + e_1$ is the peak power a server can consume. In order to achieve a power-proportional data center, workload consolidation [7], [8] is proposed. The idea of workload consolidation is to dynamically adapt the number of active servers to match the current workload so that idle servers are allowed to enter a power-saving mode (e.g., going to sleep or shutting down).

A test from VMWare [9] shows that the response time does not notably increase when the number of virtual machines (VMs) is less than six (corresponding to 80% CPU utilization), but increases rapidly after adding more VMs. To meet the QoS guarantee, c is set to the number of jobs that a server can serve without notable performance degradation. Given c , if the total number of jobs assigned to a data center is J , then the power consumption is

$$p = n * e_0 + n * e_1 + e_1 \frac{m}{c}, \quad (2)$$

where n is the integer part of $\frac{J}{c}$, and m is the fractional remainder of $\frac{J}{c}$. Equation (2) assumes that workload consolidation works perfectly such that the exact number of servers is activated to match the number of jobs assigned to the data center.

IV. ONLINE DISPATCHING ALGORITHM

The central feature of the wind-powered IDC architecture is the wind-power-aware mapping algorithm. In this section, we introduce a simple on-line algorithm that instantaneously maps incoming requests to a data center according to the current power consumption and wind power generation. The algorithm achieves two design goals: 1) it is an on-line algorithm and hence adds no extra delay to the overall response time, and 2) it exploits wind power diversity so as to maximize the wind power utilization.

A client request is first handled by a front-end load dispatching server that maps the request to one of the distributed data centers. Let S denote the set of data centers. We assume that two vectors of information are instantaneously available at the load dispatching server: 1) wind power at each wind farm $\mathbf{W}(t) = \{w_1(t), \dots, w_i(t), \dots, w_{|S|}(t)\}$, and 2) the number of remaining jobs at each data center $\mathbf{J}(t) = \{J_1(t), \dots, J_i(t), \dots, J_{|S|}(t)\}$. When the load dispatching server receives a request, it immediately selects a data center for the request based on the information. We propose a wind power-aware (WPA) data center selection algorithm that:

- 1) Computes the instantaneous power consumption of each data center according to $\mathbf{J}(t)$ by Eq. (2). We denote the power consumption vector by $\mathbf{p}(t) = \{p_1(t), \dots, p_i(t), \dots, p_{|S|}(t)\}$.
- 2) Trims wind power $w_i(t)$ by the maximum power that data center i can consume. The trimmed powers represent the effective power from the wind farm, and we denote them by a vector $\hat{\mathbf{p}}(t) = \{\hat{p}_1(t), \dots, \hat{p}_i(t), \dots, \hat{p}_{|S|}(t)\}$. (We will describe how to trim wind power in Section V-B.)
- 3) Selects the data center with minimum value of $r_i(t)$, where $r_i(t) := p_i(t)/\hat{p}_i(t)$.

In other words, the WPA data center selection algorithm first estimates the current power consumption of a data center. Then, it compares the current power consumption against the maximum wind power the data center can use to obtain the instant wind power utilization r . All the utilizations are computed and the data center with the minimum instantaneous utilization is selected.

V. WIND FARM SELECTION

The idea of wind-powered distributed IDCs is to place data centers at locations which are close to wind farms. However, the total number of servers is finite and only a few locations can be selected to build data centers. This section answers the following questions: how many data centers are needed and where should they be placed? In other words, which characteristic of wind farms should we consider when selecting locations at which to build wind-powered data centers?

A. Brownout Probability and Brown Energy Utilization

We propose two criteria to evaluate selected wind farms: brownout probability and brown energy utilization.

Let us first define the notation for our analysis: the wind power generated from wind farm i is a time series $\psi_i := (w_i(t))_{t \in \{1, \dots, T\}}$, where T is a long time interval of interest. Denote s as a subset of the wind farm set S .

Let $B(P_{max}(M), \alpha)$ denote *Brownout Probability* for M servers, with at least $\alpha P_{max}(M)$ ($0 \leq \alpha \leq 1$) power provision guaranteed to IDCs at all times regardless of the availability of wind. Because wind power is not always available, we call the state when the wind power drops below $\alpha P_{max}(M)$ as the “brownout” state. The Brownout Probability is the probability

that the system is in a “brownout” state. Formally, we have

$$B(P_{max}(M), \alpha) = \frac{1}{T} \sum_{t=1}^T I\{w[t] < \alpha P_{max}\}, \quad (3)$$

where $w(t)$ is the wind power, and the indicator function $I\{A\}$ is equal to 1 if A is true, and 0 otherwise.

Similarly, we define the *Brown Energy Utilization* $U(P_{max}(M), \alpha)$ as the percentage of brown energy in the total energy portfolio. When wind power is not enough to guarantee $\alpha P_{max}(M)$, the IDCs will draw “brown” energy from the electricity grid. The amount of “brown” power it needs to draw is the difference between $\alpha P_{max}(M)$ and current wind power generation. Thus, we compute the Brown Energy Utilization as

$$U(P_{max}(M), \alpha) = \frac{\sum_{t=1}^T \max\{\alpha P_{max}(M) - w[t], 0\}}{T \alpha P_{max}(M)}. \quad (4)$$

The numerator is the brown energy that is needed to be drawn from the grid in order to keep the power provision always above $\alpha P_{max}(M)$.

B. The Number of Servers in Each Data Center

Note that the maximum power consumption is determined by the number of servers in the data center. Thus, in order to calculate the effective wind power, we must determine how many servers there are to be at each data center. We consider a very simple way to split the servers. Assume an IDC system consisting of M servers and suppose that these servers are allocated between the wind farms. We assume that the split ratio is proportional to the average power output of the corresponding wind farm.

Therefore, the *effective wind power* at data center i at a given time t is

$$Pe_i(t) = \min\{P_{max}(M_i), w_i(t)\}. \quad (5)$$

The effective wind power means that a data center can use only up to $Pe_i(t)$ of wind power regardless of whether the actual wind power available may exceed it.

C. The Number of Data centers

To select wind farms, the first question is how many data centers should be built to achieve the best performance?

To quantify the impact of the number of data centers, we evaluate both $U(P_{max}(M), \alpha)$ and $B(P_{max}(M), \alpha)$ for different numbers of wind farms. Set $P_{max}=200$ MW, and $\alpha = 30\%$, 50% , or 70% respectively.

The results are plotted in Figures 2 and 3. Note that the results are for 1 to 5, 44, and 45 wind farms. At this stage, we have not yet discussed how to select wind farms. Thus, we use a brute force approach of evaluating all possible combinations and plot the minimum value for each fixed number of wind farms. We obtain two observations from the results: 1) spreading servers to multiple wind farms can reduce brown energy utilization by 89.4%-97.8% and brownout probability by 61.4%-95.7%; and 2) building only a few data centers (e.g., four or five) is enough to realize most (e.g., 74%-96%) of the performance gain.

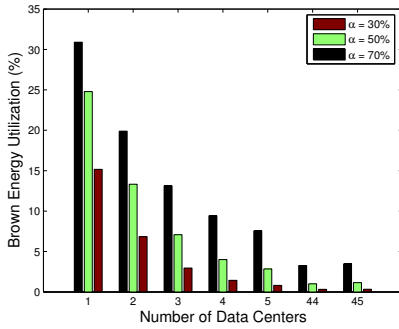


Fig. 2. Brown Energy Utilization vs. Number of Data Centers, $P_{max} = 200MW$, $\alpha = 30\%, 50\%, 70\%$

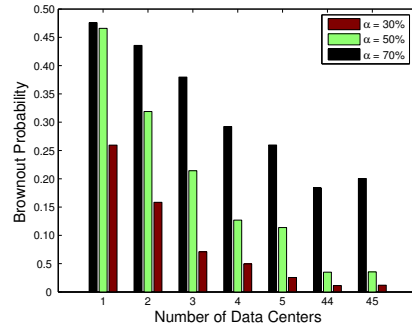


Fig. 3. Brownout Probability vs. Number of Data Centers, $P_{max} = 200MW$, $\alpha = 30\%, 50\%, 70\%$

D. Correlation Coefficient

The next question we want to answer is that if multiple wind farms are jointly considered, what property across wind farms effects the wind power utilization the most?

Intuitively, it is believed that the correlation coefficient between wind farms is the important metric to consider for selecting wind farms because well-correlated wind farms do not have diversity to exploit. Surprisingly, however, we find that the correlation coefficient is of little help in the wind farm selection. Let us look at the case of two wind farms. We compute the correlation coefficient for each pair of wind farms and then sort them in ascending order. We compute $B(200, 50\%)$ and $U(200, 50\%)$ for each pair. The result is shown in Figure 4. One can see that neither brownout probability nor brown energy utilization exhibits any clear trend with correlation coefficient.

The explanation is as follows. Combining two uncorrelated time series can help to reduce the overall variability only if the means of the two time series are comparable. Otherwise, the overall variability is determined by the time series with the larger mean. However, the power generations of wind farms could be very different. Therefore, the correlation coefficient alone is not a sufficient metric for wind farm selection.

E. Coefficient of Variation

In statistics, the *coefficient of variation* (CV) is a good metric to measure the extent of variability. It is defined as the ratio of the standard deviation σ to the mean μ . We propose to use CV as a metric to select wind farms. We pair wind farms in all possible combinations and compute brownout probability and brown energy utilization, with $P_{max} = 200$ MW and $\alpha = 50\%$ for every pair of wind farms. The result is plotted in ascending order of CV. The results are shown in Figure 5. One can see that brown energy utilization and brownout probability clearly rise as the CV increases, in both cases. These trends are fitted on a log scale. A smaller CV leads to smaller B and U . These results suggest that CV is indeed an effective metric to select wind farms for wind-powered IDCs.

VI. SIMULATION STUDAY

The simulation is based on real-world traces for both wind power and workloads. To make the problem meaningful, we assume very large IDCs that demand up to 100 MW of power at peak load. Accordingly, we scale the workload such that the peak load matches the peak power consumption.

Since the workload traces span only a period of one month, we select the same length of time for the wind power traces, e.g., the month of April 2009.

A. Effectiveness of WPA Dispatching Algorithm

We first evaluate the performance of the WPA dispatching algorithm. The baseline for comparison is a naive load balancing policy that simply balances workload across data centers regardless of the availability of wind power. In this case study, we assume that there are three data centers at different locations. Two simulations are carried out: one adopting the WPA algorithm and the other using the naive load balancing algorithm. Three wind farms were selected for the simulation. In addition to brown energy utilization, we also calculate the average data center utilization. This is to show the extent of load imbalance if IDCs use the WPA algorithm. The result is shown in Figure 6. In this figure, the solid bars represent overall brown energy utilization and the blank bars represent the average utilizations of the three data centers. With the WPA algorithm, the wind-powered IDCs reduce brown energy utilization to 16%. In contrast, in the same situation, the naive load balancer needs about 30% brown energy. As a trade-off, the naive load balancer achieves perfect balance in terms of data center utilization whereas WPA leads to unbalanced workload allocation. However, the imbalance factor is not large: the maximum is 28%, and the minimum is 20%.

B. Impact of Correlation of Variation

Next, we quantify the impact of coefficient of variation (CV) on wind-power utilization. For these simulations, we fix the number of wind farms to be three. Then, we calculate the CVs for the aggregate wind powers of all possible combinations of three wind farms. Instead of conducting simulations for every combination, we choose five groups of wind farms with respect to five ranges of CVs: [0.3,0.4], [0.5,0.6], [0.7,0.8], [0.9,1.0],

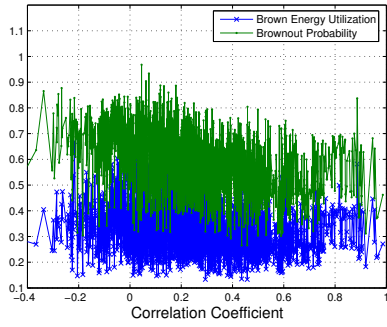


Fig. 4. Brown Energy Utilization and Brownout Probability vs. Correlation Coefficient, $P_{max} = 200MW$, $\alpha = 50\%$, All Pairs

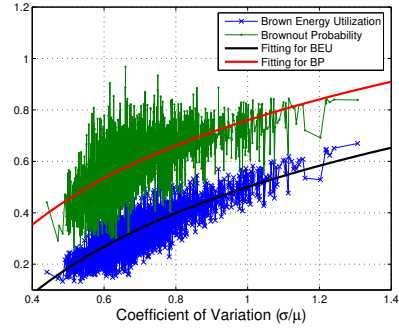


Fig. 5. Brown Energy Utilization and Brownout Probability vs. Coefficient of Variation, $P_{max} = 200MW$, $\alpha = 50\%$, All Pairs

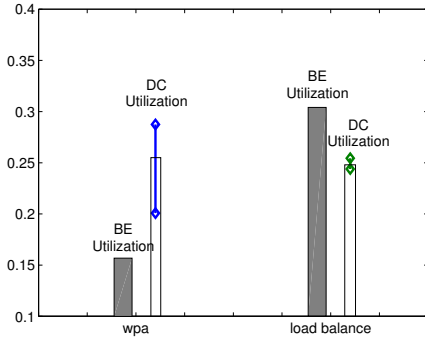


Fig. 6. Brown Energy (BE) Utilization and Data Center (DC) Utilization: WPA vs. naive load balance.

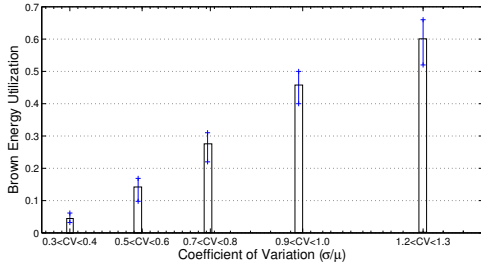


Fig. 7. Brown Energy Utilization vs. Coefficient of Variance

and [1.2,1.3]. Then, we select randomly five wind farm combinations from each group and perform the simulations for them. The results are presented in Figure 7. In this figure, the mean, the maximum and the minimum brown energy utilizations of the five groups are plotted, respectively. One can observe an obvious increase of brown energy utilization with increase of CV. We find that a good selection of wind farms can significantly improve the wind power utilization. In particular, the simulation shows that a good selection results within as little as 3.4% brown energy utilization (i.e., 96.6% energy from wind), whereas a bad selection causes IDCs to purchase 66% energy from the grid (i.e., only 34% energy from wind).

VII. CONCLUSION

Through the analysis and trace-based experiments, we have made a case that it is actually possible for a distributed IDC system to exploit the geographical and temporal diversity of wind power to nearly achieve “entirely green” cloud-scale services. We have demonstrated the following as keys to increasing wind power utilization: 1) The WPA policy can efficiently allocate workloads to data centers where on-site wind power is available; 2) Wind farm selection is an important factor; and 3) The coefficient of variation is an effective metric to evaluate the selection of wind farms. Our workload and wind power models are based on actual data. We do make a number of simplifying assumptions that perhaps effect the actual effectiveness of the wind power. However, many emerging technologies such as online migration across data centers, large scale energy storage solutions and renewable-aware scheduling have not been explored in this paper. Thus, we argue that powering data centers entirely with wind energy should be feasible.

VIII. ACKNOWLEDGMENT

Special thanks to Dr. Chen Zhang from McGill University for his help with workload trace analysis. His great effort made the simulation study faster and easier.

REFERENCES

- [1] C. Miller, “Solar-powered data center,” *Data Center Knowledge*, 2010.
- [2] —, “Wind-powered data centers,” *Data Center Knowledge*, 2010.
- [3] P. Thibodeau, “Wind power data center project planned in urban area,” *ComputerWorld*, 2008.
- [4] “Green house data: Greening the data center,” <http://www.greenhousedata.com>.
- [5] C. Stewart and K. Shen, “Some joules are more precious than others: Managing renewable energy in the datacenter,” *Workshop on Power Aware Computing and Systems (HotPower)*, 2009.
- [6] L. Barroso and U. Holze, “The case for energy-proportional computing,” *IEEE Computer*, 2007.
- [7] A. Whitaker, M. Shaw, and S. D. Gribble, “Lightweight virtual machines for distributed and networked applications,” *Technical Report 02-02-01*, University of Washington, 2002.
- [8] P. Padala, K. G. Shin, X. Zhu, M. Uysal, Z. Wang, S. Singhal, A. Merchant, and K. Salem, “Adaptive control of virtualized resources in utility computing environments,” *The 2nd ACM SIGOPS/EuroSys European Conference on Computer Systems*, 2007.
- [9] VMWare, “Performance study: SQL server workload consolidation,” *VMWare*, 2008.