Renewable Energy Prediction for Improved Utilization and Efficiency in Datacenters and Backbone Networks

Baris Aksanli, Jagannathan Venkatesh, Inder Monga and Tajana Simunic Rosing

Abstract Datacenters are one of the important global energy consumers and carbon producers. However, their tight service level requirements prevent easy integration with highly variable renewable energy sources. Short-term green energy prediction can mitigate this variability. In this work, we first explore the existing short-term solar and wind energy prediction methods, and then leverage prediction to allocate and migrate workloads across geographically distributed datacenters to reduce brown energy consumption costs. Unlike previous works, we also study the impact of wide area networks (WAN) on datacenters, and investigate the use of green energy prediction to power WANs. Finally, we present two different studies connecting datacenters and WANs: the case where datacenter operators own and manage their WAN and the case where datacenters lease networks from WAN providers. The results show that prediction enables up to 90% green energy utilization, a 3x improvement over the existing methods. The cost minimization algorithm reduces expenses by up to 16% and increases performance by 27% when migrating workloads across datacenters. Furthermore, the savings increase up to 30% compared with no migration when servers are made energy-proportional. Finally, in the case of leasing the WAN, energy proportionality in routers can in-crease the profit of network providers by 1.6x.

Baris Aksanli
University of California, San Diego. La Jolla, CA, USA.e-mail: baksanli@ucsd.edu

Jagannathan Venkatesh
University of California, San Diego. La Jolla, CA, USA. e-mail: jvenkate@ucsd.edu

Inder Monga
Energy Sciences Network, Lawrence Berkeley National Laboratory. Berkeley, CA, USA. e-mail: imonga@es.net

Tajana Simunic Rosing
University of California, San Diego. La Jolla, CA, USA. e-mail: tajana@ucsd.edu

1 Introduction

As the demand for computing increases globally, the number of datacenters has increased to meet the need. Recent studies indicate that the total power consumption of all datacenters in the world has increased by 56% from 2005 to 2010 [25], with associated global carbon emissions and an estimated annual growth rate of 11%. Their energy needs are supplied mainly by non-renewable, or brown energy sources, which are increasingly expensive as a result of availability and the introduction of carbon emissions taxes [37]. Consequently, several datacenter operators have turned to renewable energy to offset the energy cost.

The integration of renewable energy is complicated by the inherent variability of its output. Output inconsistency typically leads to inefficiency due to lack of availability or sub-optimal proportioning, which carries an associated financial cost. These costs are mitigated in various ways: several datacenter owners, such as Emerson Networks, AISO.net, and Sonoma Mountain Data Center supplement their solar arrays with utility power, and other datacenter owners, such as Baronyx Corporation and Other World Corporation, have been forced to augment their input power with other forms of energy or through over-provisioning, respectively [19]. Previous investigation into the existing methods in datacenter green energy demonstrates that variability results in low utilization, on average 54%, of the available renewable energy [38].

A number of publications investigated the best strategy to manage renewable energy as a part of datacenter operation. The work in [19] reduces the peak datacenter power with local renewable sources and power management algorithms. They investigate power capping, both of individual servers using dynamic frequency scaling, and of server pools by reducing the number of machines utilized in each pool. However, they have significant quality-of-service (QoS) violations when limiting peak power. The study in [38] explores brown energy capping in datacenters, motivated by carbon limits in cities such as Kyoto. The authors leverage distributed Internet services to schedule workloads based on electricity prices or green energy availability. By defining workload distribution as a local optimization problem, the authors demonstrated 35% lower brown energy consumption with a nominal (10%) hit on service level agreement (SLA) violations. Similarly, [40] optimizes for energy prices, to reduce overall energy consumption by distributing workloads to datacenters with the lowest current energy prices. The insight is that renewable sources such as solar energy are actually cheapest during the day, when workloads are at the highest and utility sources are most expensive. Job migration is then modeled as an optimization problem, and the authors identify a local minimum energy cost among the available datacenters that still meets deadlines. The results demonstrate that their algorithm performs within 5.7% of the optimum distribution, a significant improvement over established greedy algorithms. The authors of [29] analyze the opportunities and problems of using supply-following loads to match green energy availability. When green energy is insufficient, workloads are terminated or suspended, restarting or resuming when availability returns. However, the results show very low green energy efficiency and a failure to meet required service-level guarantees.

The above datacenter examples demonstrate the high cost and necessary precautions needed to successfully use highly variable green energy, at the cost of efficient utilization. However, an important means of reducing such variability remains overlooked in datacenters: green energy prediction. In [6], we investigated the existing methods in solar and wind energy prediction, developing prediction algorithms suitable for the datacenter domain. We implemented and evaluated our algorithms in a case study, leveraging prediction to improve green energy utilization in datacenters by 90%, a 3x improvement over the existing methods.

Previous publications concerned with energy costs primarily propose a follow the sun cost-management strategy [19], [29], [32], [8], [33], [3] and generally neglect the cost of wide area net-working (WAN) incurred by job migration between datacenters. This assumption is reasonable for small datacenter networks that own the WAN and incur low network costs. Consequently, related work has WANs used to increase system performance via load balancing [51], [36], [34] or improve energy efficiency by migrating jobs [8], [33], [3]. However, these arguments are not applicable for large WAN costs and datacenters that lease the network.

Datacenters lease the WAN by agreeing to pay a certain price for a fixed bandwidth usage. However, as WAN usage increases, network owners [15], [5] offer Bandwidth-on-Demand services, especially for r-datacenter applications [1]. Additionally, the WAN may take up to 40% of the total IT energy cost, and is expected to continue growing as demand for distributed data processing continues to rise [17] and as the server hardware becomes more energy efficient [11]. With the increasing importance of managing energy consumption in the network, WAN providers can charge users not just on the amount of bandwidth they use, but also the time of day when they use it. For example, using the network in a peak hour may be more expensive than when it is idle, reflecting electricity market prices [50]. Additionally, with the introduction of carbon taxes, WAN providers can also vary energy prices depending on the energy source. Consequently, datacenters might be open to longer, less expensive paths on the network. For example, a datacenter may request a path that uses green energy to avoid paying extra carbon emission taxes, or a less-utilized path to avoid extra utilization costs. Our work uniquely considers both the costs of geographically distributed datacenters and the profits of the network provider. We analyze different network cost functions, along with the analysis of new technologies that would allow using more energy proportional routers in the future.

In this chapter, we first evaluate the advantages of short-term green energy prediction on the datacenter scale. We explore the existing in short-term solar and wind energy prediction methods, applying each to real power traces to analyze the accuracy. Using green energy prediction in local renewable energy sites and varying brown energy prices, we propose an online job migration algorithm among datacenters to reduce the overall cost of energy. While such job migration has been studied extensively before, we uniquely consider network constraints such as availability, link capacity and transfer delay at the same time. By including these constraints in our framework, we model the impact of the network and create a more holistic multi-datacenter model. Additionally, we investigate the impact of two aspects of datacenter operation typically overlooked in previous work: tiered power pricing, which penalize the datacenter for exceeding certain level of power restrictions with as much as 5x higher energy costs [28], and WAN leasing costs/cost models, which leverage energy-aware routing. Both play a significant impact in datacenter job scheduling, reflected in our results.

We also analyze the impact of new technologies in datacenter WAN, such as energy-proportional routing, green energy aware routing, and analyze leasing vs. owning the WAN. Our work is the first analyzing different WAN properties in a job migration algorithm involving both mixed energy sources and prediction. We observe that green energy prediction helps significantly increase the efficiency of energy usage and enables network provisioning in a more cost effective way. Similarly, we show that using a WAN to transfer workloads between datacenters increases the performance of batch jobs up to 27% with our performance maximization algorithm, and decreases the cost of energy by 30% compared to no data migration with our cost minimization algorithm. Unlike previous works, we show the potential for green energy to go beyond simply cost reduction to improving performance as well. Our analysis of leasing WAN shows that network providers can increase profits by charging datacenter owners by bandwidth, but datacenters can still benefit by using dynamic routing policies to decrease their energy costs. Furthermore, as servers and routers become more energy proportional, we demonstrate increases in both datacenter cost savings and network provider profits.

2 Green Energy Prediction and Datacenters

2.1 Solar energy prediction

Solar energy algorithms exploit the daily pattern of solar irradiance, a primary factor in determining power output. The simplest algorithms are based on exponential weighted moving average (EWMA) [23]. Several extensions to the EWMA algorithm have been proposed, incorporating multiple days' predictions to derive a more representative average value in variable weather [12], [43]. Extended EWMA, eEWMA, [12] uses previous days' measurements to account for the error of each slot. The weather-conditioned moving average (WCMA) algorithm [43] takes into account the actual values from previous D days and the current day's previous N measurement slots. It averages the values for the predicted slot from previous days and scales it with a GAP factor, which represents the correlation of the current day against the previous days:

$$E(d,n+1) = \alpha.E(d,n) + GAP_k.(1-\alpha).M_D(d,n+1) \quad (1)$$

where $M_D(d,n+1)$ represents the median of previous days' values, and GAP_k represents the scaling of the current day against the previous days. The inclusion of both patterns from multiple previous days as well as the use of values from the current day itself help WCMA provide a better pattern for the performance of solar panels. The three algorithms discussed above are tested using real solar power traces from the UCSD Microgrid. Absolute mean error is calculated against the measured data, shown in Table 1. The optimal parameter values have been determined empirically for each algorithm. The results demonstrate the importance of incorporating recent data to reduce error. The WCMA algorithm provides a significant improvement over EWMA and extended EWMA algorithms due to its highly adaptive nature, and its effective use of the GAP factor to scale future intervals based on the deviation from previous days.

Table 1 Solar Power Prediction Algorithm Comparison

Algorithm	**Absolute Mean Error (%)**		
	Consistent Conditions	**Inconsistent Conditions**	**Severely Inconsistent Conditions**
EWMA	12.7	32.5	46.8
eEWMA	4.9	23.4	58.7
WCMA	4	9.6	18.3

2.2 *Wind energy prediction*

Wind prediction algorithms may use physical or statistical models. Physical models use spatial equations for wind speed at the locations of each turbine, and then predict wind power with theoretical or measured power curves [10]. Statistical models aggregate measured or forecasted meteorological variables and develop a relationship between the variables and the output power.

Several data-mining models have been used to predict the wind speed based on the meteorological variables collected from SCADA data acquisition units at each wind turbine [30]. The heuristics developed for wind speed prediction are then applied to wind power prediction, demonstrating 19.8% mean error for 10-minute-ahead prediction. Nearest-neighbor tables (k-NN) algorithm reduces this error to 4.23% by mapping wind speed to wind power [31]. However, when forecasted wind speed is used, the power prediction error grows to 27.83%.

Power curves, which describe the output power of wind turbines mapped against wind speed, form the basis of many predictors [10], [16], [30]. The work presented in [46] analyzes power curves and demonstrates their inaccuracy. Instead, the paper uses a dynamic combination of several statistical predictors, most notably the Auto-regressive Moving Average (ARMA) model with past wind power, speed and direction as inputs. The results show a 50% reduction in power prediction error, with ability to reduce error levels between prediction horizons of 2 to 45 hours.

The above algorithms pose difficulties in implementation: unlike the solar prediction algorithms, which only require past power data, the wind prediction algorithms require various types of high-overhead input. Instead, we dramatically lower the overhead with our new wind energy predictor: we construct weighted nearest-neighbor tables based on the two most correlated variables contributing to wind energy output: the wind speed and direction [6]. The weighted tables show preference to the most recent results and allow the algorithm to adapt to gradual changes, while the power curves, based on both wind speed and direction, provide versatility. The algorithm to add a new entry to the table is in equation 2, where $P_{new}(v,d)$ is the new power curve table entry for a given wind velocity v and direction d, $P_{old}(v,d)$ is the existing value, and $P_{obs}(v,d,t)$ is the observed value at time t. Future interval prediction is determined by equation 3.

$$P_{new}(v,d) = \alpha.P_{obs}(v,d,t) + (1-\alpha).P_{old}(v,d) \tag{2}$$

$$P_{pred}(v,d,t+k) = P(v(t+k),d(t+k)) \tag{3}$$

The algorithms described above have been tested using real power data from a Lake Benton, MN wind farm, and the meteorological data was provided by published reports from the National Renewable Energy Laboratory (NREL). For better comparison, we have all the predictors use the same inputs: wind speed and direction. We also include the commonly used baseline - persistence prediction, which assumes that the future interval is the same as the current one.

Persistence has a high error at 137%, affected by the high variability of the wind farm power. The data-mining algorithm's error is at 84%, despite using the two most-correlated variables. This can be attributed to the unreliability of using forecasted wind speed for a region as opposed to measured wind speeds at each turbine level. The ARMA model performed better, at 63% error, but the accuracy is hampered by the limited input data available. The wind-speed-based nearest-neighbor predictor, the kNN algorithm, performed the best, with an error of 48%, which can be attributed to the higher variance of the Lake Benton wind farm than the wind farm in the original work. Our custom nearest-neighbor predictor, which uses both wind speed and direction to develop a relationship with wind farm energy output, has only 21% mean error as it is more adaptive to recent conditions and as a result is 25% better than the next-best algorithm.

Table 2 Wind Power Prediction Algorithm Comparison

Algorithm	***Mean Error (%)***	***Std. Dev. (%)***
Persistence	137	340
Data-mining [30]	84	101
ARMA Model [46]	63	12
kNN Predictor [29]	48	32
Our NN predictor [6]	21	17

3 Datacenter and WAN Models

Multiple datacenters increase the capacity, redundancy, and parallelism of computation, but a fast, reliable inter-datacenter network is essential to maintain performance. Since large datacenters consume a lot of power, they usually undergo a tiered power pricing. The tier level depends on the overprovisioned input power to avoid high prices in-peak periods [28]. This can be seen as a power budget. In this work, we also study the effects of different power tier levels and how these levels can affect the job migration decisions and network utilization. To avoid power tier violations, datacenters may deploy several techniques: CPU capping, virtual CPU management and dynamic voltage and frequency scaling (DVFS), all of which incur performance overhead. If the power goes beyond the tier level, it is charged at higher rates, which can be 3x10x higher than the base cost [40], [3]. One way to remedy this problem is to leverage a network of globally distributed datacenters along with renewable energy prediction for peak power reduction. In the next subsections, we present our datacenter and backbone network models, which we then use in our frameworks for managing renewable-powered globally distributed datacenters and related WAN.

3.1 Backbone network model

Our network topology is a subset of the LBNL ESnet [2], containing 5 datacenters and 12 routers in total, distributed over the USA (figure 1), where each link has a predefined capacity from 10Gbps to 100Gbps. A portion of this capacity is normally reserved for background traffic (10% in our experiments). When calculating the network energy cost, we consider the router power and add a fixed offset for the link power consumption. The power consumption of the router can be estimated using a simple linear model [44] based on bandwidth utilization.

Current router's idle power is very close to active power. In our previous work, we investigate methods for designing more energy-proportional routers and construct

Fig. 1 Network Topology; squares=datacenters, circles = routers

representative power proportionality models [2]. Figure 2 reflects our findings: the non-proportional curve represents an actual state-of-the-art router [47], and the step function depicts shutting down idle line cards. Smooth proportionality is a linear correlation of the step function, while ideal proportionality reflects an idle router with no power cost.

In our model, we account for the network transfer delay as an increase in the total response time of a job. The state-of-the-art is Dijkstras Shortest Path Routing (SPR). For comparison, we use our own Green Energy Aware Routing algorithm (GEAR), which minimizes the amount of brown energy used [2]. GEAR is a power- and performance-constrained dynamic routing algorithm that performs online path computation and constructs virtual circuits [21]. We use different price models to calculate the network lease costs including, fixed bandwidth (BW) cost, where cost does not increase with utilization; and linear BW cost increase, which assumes that cost of operation and revenue are proportional to usage. These two options represent the different models that network operators might incorporate in their service level agreements (SLAs). In the results section, we show how the total cost is affected by these cost schemes.

3.2 Datacenter model

In order to represent a multi-datacenter network more accurately, each datacenter is modeled separately based on actual measurements. Each includes computation models to represent servers and how they interact with each other and the workloads they execute. For each datacenter in Figure 1, we implement a detailed model designed to faithfully represent a fully populated set of datacenter containers. Each container has 8 racks populated with 250W Intel Xeon Quad core machines [24]. We create and run multiple virtual machines using a Xen VM [18], and measure runtime characteristics such as CPU, memory, and power consumption. We use these measurements to construct our datacenter model for simulation.

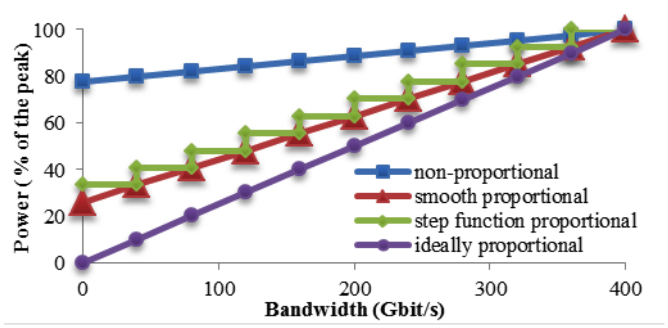

Fig. 2 Router energy proportionality schemes

The workload is divided into two representative categories: service and batch jobs, both of which are run on our servers. We model the former with RUBiS [45] and the latter with MapReduce [22]. Short-running service jobs have tight latency constraints, and the longer-running batch jobs represent the background tasks (i.e. indexing results or correlating data) and need to maintain a certain level of throughput. Their inter-arrival time distributions are modeled with a lognormal distribution based on measured values [6]. We model service and batch job arrival to the system independently, and place them in separate local job queues. Each MapReduce job consists of multiple tasks and multiple tasks are dispatched from an active job and put in different servers.

Servers run the jobs assigned to them, and prioritize service jobs over batch jobs because of their response time requirements. These requirements are indicators for the quality of service (QoS) a datacenter has to maintain to ensure its profitability. We also measure the interference of running different types of jobs simultaneously on the same server. Since these jobs have different resource requirements, the interference of one on the other might lead to performance degradation for either job. In our experiments, we observe that service requests have negligible impact on MapReduce jobs, but MapReduce jobs are detrimental to both service jobs and other MapReduce jobs. In order to meet QoS of service jobs and maintain the throughput hit of batch jobs to fewer than 10%, we limit the total number of simultaneous MapReduce jobs on a single server. The baseline of this study is established in [6].

We calculate the server power consumption with a linear CPU-utilization based equation [13] and scale the aggregate server power cost using power usage effectiveness (PUE) metric, which is set to 1.15 [20], to find the total power of the datacenter as a function of overheads related to cooling, powering, and other loads [7]. The deviation between our simulations and measurements is 3% for power consumption, 6% for service job QoS and 8% for MapReduce job performance.

4 Relocating Jobs to Improve Efficiency

4.1 Background

Multi-datacenter networks offer advantages for improving both performance and energy. As each datacenter is in a different location, its peak hours and energy prices vary. A datacenter with high electricity prices may need to migrate work to another datacenter with a lower price, incurring some performance and power cost due to data migration. The live migration of virtual machines over high speed WAN has made this idea feasible, as it offers fast transmission with limited performance hit [14]. However, the migration costs through WAN need to be considered. For example, WAN may be leased, with lease costs quantified per increment of data transferred, and thus might be too high to justify frequent migration of jobs between datacenters [4]. Furthermore, datacenters often undergo a tiered power pricing scheme. The energy under a specific level may cost a fixed amount and this fixed price changes depending on the location, so it is beneficial to run jobs in a datacenter at a lower fixed price. Data migration should not increase the total power consumption to more than the amount specified by the specific tier level. Otherwise, extra power costs are calculated using higher prices, generally much higher than the fixed price.

Table 3 Summary and comparison of the related work

	Buchbinder 2011 [40]	**Qureshi 2009 [4]**	**Mohsenian-Rad 2010 [39]**	**Rao 2010 [33]**
Goal	Minimize electricity bill	Minimize electricity bill	Minimize carbon footprint & job latency	Minimize electricity bill
How	Move jobs where energy is cheaper	Move jobs where energy is cheaper	Migrate jobs to different locations depending on the goal	Move jobs where energy is cheaper
Workload	No specification	No specification	Service requests only	Service requests only
Perf. Constraints	X	X	Latency of service requests	Latency of service requests
Network Cost Model	Fixed cost per bandwidth	Fixed cost per bandwidth	X	X
Routing	X	Distance based routing	X	X
Green Energy	X	X	Local green energy, carbon tax	X
Network Delay	X	✓	✓	X

	Liu 2011 [51]	**Le 2010 [26]**	**Aksanli 2011 - 12 [6], [2]**
Goal	Minimize brown energy use	Minimize the total cost of energy	Maximize batch job performance & Minimize brown energy use
How	Move jobs to local green energy	Forward jobs to datacenters	Move jobs where utilization is low
Workload	No specification	Different job types (not explicitly specified)	Mix of service and batch jobs
Perf. Constraints	X	SLA of service requests	Latency of service requests and throughput of batch jobs
Network Cost Model	X	X	X
Routing	X	X	Static routing vs. energy aware
Green Energy	Local Green Energy	Grid green energy carbon tax	Local green energy with prediction
Network Delay	X	X	✓

Table 3 summarizes and compares the key state of the art contributions for managing distributed datacenters in order to minimize an objective function, e.g. the overall cost of energy. Buchbinder et al. [40], Qureshi et al. [4] and Rao et al. [33] relocate jobs to where the energy is cheaper to minimize the energy cost. They do not model different energy types; perform detailed workload performance analysis and different routing options for both WAN providers and datacenters. Le et al. [26] solves a similar problem including green energy in their model but they assume a centralized dispatcher and do not analyze network latency or cost. Liu et al. [51] and Mohsenian-Rad et al. [39] minimize the brown energy usage or carbon footprint. They either do not consider the variability of green energy or do not have a network model. Aksanli et al. [2] solve a load-balancing problem by modeling network properties, but do not consider energy costs. As we can see from this analysis, previous studies do not consider all the important aspects of multi-datacenter networks simultaneously in their models. As we show in this chapter, this can lead to overestimated cost savings or overlooked performance implications due to not considering both the requirements of different types of applications and WAN characteristics.

In this chapter, we generalize the problem of migrating jobs among datacenters to minimize the cost of energy and analyze the effects of using WAN for the transfer. Our design considers both brown and locally generated green energy, and variable energy market pricing. We simultaneously investigate energy proportionality of routers and servers and tiered energy pricing, which are at best considered individually in previous works. Additionally, we account for the latency and cost of the WAN, the costs of leasing or owing the WAN, and the impact of different routing algorithms. Our work is also the first showing the potential of green energy to improve performance in addition to addressing environmental concerns or reducing energy costs.

4.2 Cost minimization and performance maximization algorithms

We now describe our cost minimization algorithm, which considers the properties of both the datacenters and the backbone network simultaneously. Our algorithm performs in discrete time steps of 30 minutes. Each datacenter has its own workload distributions that represent different types of applications in a datacenter environment. The properties of these distributions are determined by applying statistical analysis on real datacenter traces (section 3.2 outlines the distributions and section 5 presents the real workloads we use).

The goal of our algorithm is to determine which workloads we need to transfer among different datacenters during each interval to minimize the energy cost. The current algorithm assumes a centralized implementation for control for job migration decisions, though each datacenter generates its own workloads. We assume that green energy is generated and used locally, and is prioritized over brown energy to minimize the total cost, as green energy is a fixed, amortized cost. Thus, we transfer workloads to datacenters which have available capacity and extra green energy. Because of datacenters' energy pricing scheme, energy in a particular location may have a fixed, low cost up to a specified amount of peak power capacity. After this level, energy becomes much more expensive. Therefore, our goals include maintaining utilization in datacenters such that we do not increase the power consumption further than the power tier levels.

Figure 3 illustrates our cost minimization algorithm. Each interval begins with the calculation of the amount of energy required by each datacenter, incorporating the previous and incoming load rates. The former represents the active jobs at a given time, and the latter is determined by the statistical distributions of real applications. We estimate the green energy availability using prediction (section 2), obtain the current brown energy pricing, and check power restrictions. Based on the energy need and green energy availability, each datacenter determines if it has surplus green energy. The key assumption is that if brown energy has already been within the lower price region, it makes sense to use it for running jobs, while green energy can be used to both reduce power consumption and to run extra jobs which otherwise might not be scheduled.

Then workloads are transferred from the datacenters with the highest need to those with the highest available green energy. The workload that can be transferred from a datacenter is determined by what is migrateable, while the workload that can be transferred to a particular datacenter is limited by the amount of additional green energy and WAN availability. This process continues until every datacenter is analyzed. If there are workloads remaining in any datacenters at the end, the algorithm focuses on datacenters with the cheapest brown energy cost. It moves workloads from the datacenters with higher energy costs to those with the cheapest brown energy. The amount of data that can be transferred is limited by receiving datacenter's peak power constraints and tiered power levels. If there are still unscheduled jobs remaining at the end of this process, they are scheduled in datacenters where the market electricity prices are the lowest.

We can also modify this iterative part of our algorithm to maximize the performance of the workloads instead of minimizing the total cost of energy. In this case, we transfer the jobs that are actively waiting in the execution queue to datacenters with excess green energy availability. The iterative process of the cost minimization algorithm is also valid here, but the migration depends only on green energy avail-

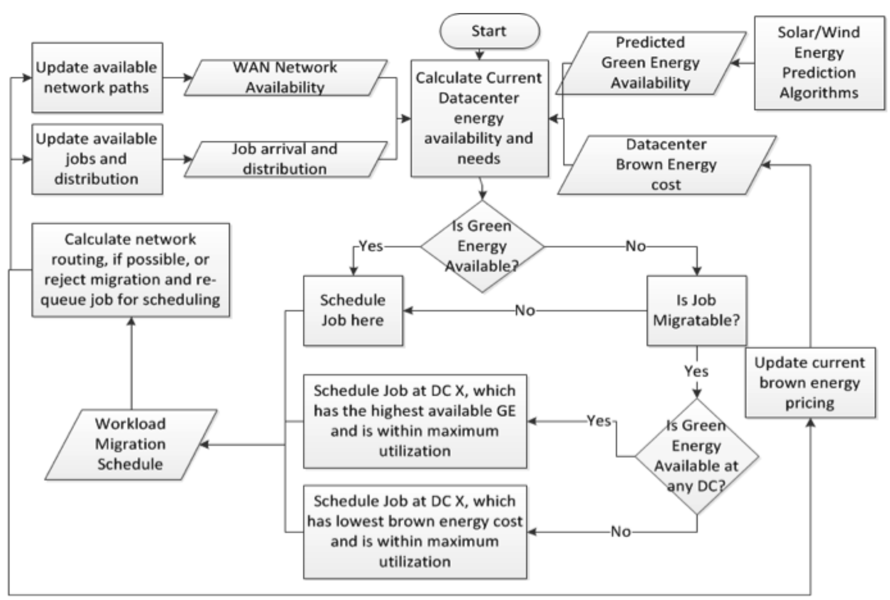

Fig. 3 High-level overview of the algorithm

ability, i.e. jobs are not migrated to datacenters with cheaper brown energy prices because extra brown energy would be required for these additional jobs. We denote this process as performance maximization as it runs additional jobs with surplus green energy.

At the end of this iterative process, we obtain a matrix representing workload transfers among datacenters. This transfer matrix is then provided to the networking algorithm, which calculates the paths to be used and the amount of bandwidth that needed by each selected path. In our study, we analyze different path selection algorithms, such as shortest path routing (SPR), green energy aware routing (GEAR), and network lease based routing. A detailed description of SPR and GEAR implementations is in [2]. Network lease based routing selects the path with the least per-bandwidth price in the case the WAN is leased. In our results, we analyze different network cost functions as well. If a selected path in the transfer matrix is unavailable due to network limitations, the job is rescheduled with a limitation on target datacenters.

Our algorithm is similar to those proposed in previous studies, but it minimizes the cost of energy more comprehensively. This is because it has a more complete view of datacenter energy costs, modeling both fixed energy costs under fixed amounts and variable, higher tier energy prices. This helps us to calculate the energy cost savings in a more accurate way. Secondly, it considers the side effects of the WAN, analyzing both the performance implications of different routing algorithms and additional leasing costs if necessary. This is key when multi-datacenter systems lease the WAN. Job migration may not be feasible for those systems if the cost of leasing the network is too high. Third, the green energy availability information is enhanced by using prediction which can provide information 30-minute ahead and thus help us allocate the workloads across multiple datacenters in a more effective manner. Last but not the least; our algorithm is flexible in the sense that it can perform for both cost minimization and performance maximization purposes. Also, we are the first to show that green energy can be used to maximize the performance rather than just minimizing the total cost of energy of geographically distributed multi-datacenter systems.

5 Methodology

We use an event-based simulation framework to analyze and compare the results of our solution to the problems described above. The inputs to our simulator are derived from measurements performed on our datacenter container (section 3.2) and data obtained from industrial deployments. This section discusses how we construct the simulation environment, including the datacenter loads, simulation parameters, green energy availability, and brown energy prices.

5.1 Datacenter load

We analyze real datacenter workload traces to accurately capture the characteristics. We use a year of traffic data from Google Orkut and Search, reported in the Google Transparency Report [20], to represent latency-centric service jobs and reproduce the waveform in Figure 3 from MapReduce [49] workloads to model throughput-oriented batch jobs. In Figure 4, we show a sample workload combination of these jobs. We use this data to find the parameters of the statistical workload models fed into our simulator (section 3.2), listed in Table 4. We also only migrate batch jobs due to the tight response time constraints of service jobs.

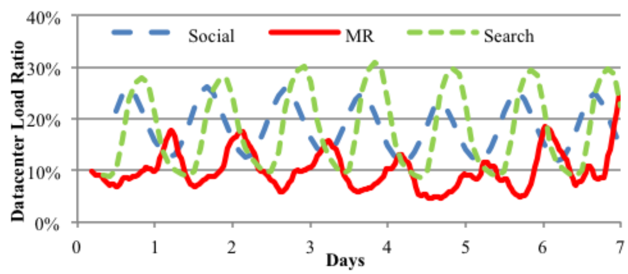

Fig. 4 Sample datacenter load ratio with different job types

5.2 Green energy availability

In Figure 5, we show a subset of the green energy availability measurements. Solar data is gathered from the UCSD Microgrid and wind data is obtained from a wind farm in Lake Benton, MN, made available by the National Renewable Energy Laboratory. The representative outputs for the other various locations in our experiments (San Francisco, Chicago, etc.) are obtained by scaling and time-shifting the measured results from our available sources to published average availability data for the target areas [41], [42].

5.3 Brown and green energy costs

Datacenters contract power from utilities to obtain competitive prices for their expected loads. This can be seen as a tiered pricing scheme. If a datacenter exceeds the tiered amount in an interval, it is relegated to higher prices, sometimes even market prices. We obtain sample fixed pricing for the midwest, the east and the west coasts

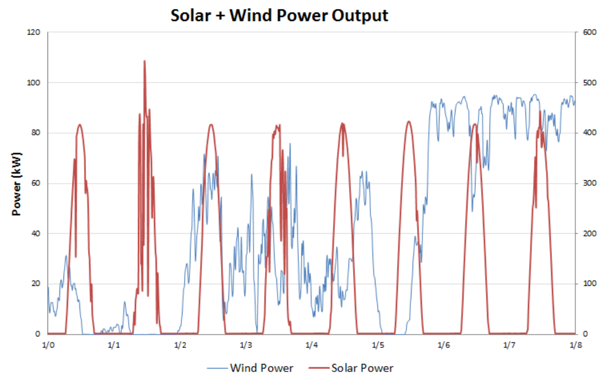

Fig. 5 Solar and wind energy availability

[27]. Since market prices change over time, we use the California ISO [9] wholesale pricing database to obtain brown energy prices for various California locations, and time-shift and scale those values for the other locations based on published averages [48]. Figure 6 shows daily pricing values for brown energy in comparison to fixed costs. The straight lines correspond to fix, under-tier prices and the others show samples of variable, market prices which can be used to charge datacenters that go over their tiered amounts.

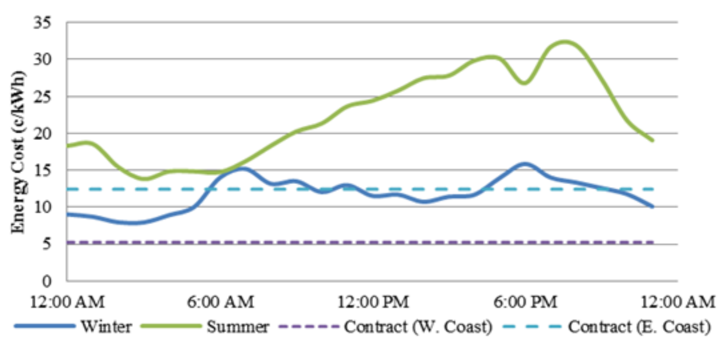

Fig. 6 Daily brown and amortized green energy cost (¢/kWh)

Local green energy costs are typically amortized over the lifetime of an installation, incorporating the capital and the maintenance costs. This is represented by a fixed offset to our cost model. We use data from [38] to obtain the capital and operational expenses of several solar and wind farms, amortized over their lifetimes, as representative solar and wind costs per interval.

We list our simulation parameters in Table 4 and present our network topology and green energy locations in Table 5. Green energy is scaled to 80% of peak datacenter and router energy needs.

Table 4 Network parameters used in simulation

Parameter	Value	Parameter	Value
Mean Web Request Inter-arrival time per client	5 ms	Average # tasks per MR job	70
Mean Web Request Service time	20 ms	Avg. throughput level per MR job	0.35
Service Request SLA	150 ms	Servers in a datacenter	1000
Mean MR Job Inter-arrival time	2 min	Number of datacenters	5
Mean MR Task Service time	4 min	Number of routers	12
Idle Server Power	212.5 W	Idle Router Power	1381 W
Peak Server Power	312.5 W	Peak Router Power	1781 W
Single link capacity	100 Gbps	Average batch VM size	8 GB

Table 5 Available renewable energy type for each location

Location	Node	Type	Location	Node	Type
Chicago	DC + router	Wind	New York	DC + router	Wind
Atlanta	DC + router	Solar	San Diego	DC + router	Solar
Kansas	Router	-	El Paso	Router	Solar
Nashville	Router	Wind	Cleveland	Router	Wind
San Francisco	DC + router	Solar & Wind	Houston	Router	Solar
Denver	Router	-	Washington DC	Router	-

6 Results

This section presents the simulation results for the base case of *no migration*, and the workload migration policies for *performance maximization* and *cost minimization*.

6.1 No migration

In this scenario, each datacenter runs its own workload using only locally available green energy. This is the baseline for our comparisons, as it represents the nominal brown energy need and quantifies the performance of batch jobs without the overhead of migration. A power tier level accounts for 85% of datacenter's power needs, while the rest, when needed, is provided at variable market prices. We allow service and batch jobs to run on the same servers while ensuring that they meet quality of service (QoS) requirements (service job $QoS_{ratio} < 1$), and find that the average MapReduce job completion time is 22.8 min. Only 59% of the total green energy supply is consumed by datacenters locally, motivating the distributed policies described previously. The next two sections quantify the impacts of *performance maximization* and *cost minimization* policies.

6.2 Performance maximization using migration

In this algorithm, we leverage migration to complete more batch jobs than previously possible. Datacenters with high utilization transfer jobs to locations with low utilization or where there is excess green energy, effectively completing more work in the same amount of time.

Most MapReduce jobs (representative of batch jobs) complete within 30 min [6], which becomes the threshold for both the green energy prediction interval and the interval for checking datacenter utilization. At each interval, the controller retrieves the resource usage and green energy profiles of each datacenter and optimizes the system by initiating extra workloads in datacenters with green energy availability while still meeting under-tier power constraints. It calculates the available transfer slots between each end-point pair, and selects the tasks to be executed remotely from each datacenter's active batch jobs. Once the tasks finish execution in a remote datacenter, the results are sent back to the original center. The key to this policy is that waiting tasks are migrated, as opposed to active tasks, resulting in more jobs executed overall (section 4).

Our simulation results show that the average completion time of MapReduce jobs is 16.8 min, 27% faster than the baseline, with no performance hit for service requests. Furthermore, since we are leveraging all available green energy for extra workloads, the percentage of green energy used is 85%, significantly higher than the baseline.

Figure 7 reports the total cost normalized against the no migration case with different tier levels specified as a percentage of the datacenter's peak power capacities and network lease options. Without tiered energy pricing (where all the consumption is charged using market prices), we demonstrate a 25% increase in the total energy cost. However, when we do include tiered energy pricing, we see more accurate results, with a cost increase of only 12% for a 70% level, and a total cost increase of 6% for an 85% level.

Since the WAN may not be owned by a datacenter, we also analyze the case where the network is leased. In this case, a bandwidth-dependent cost is incurred. Figure 7 shows the results of this analysis over different cost functions that net-work providers use. For linear increase (section 3.2), we see that the network cost can be up to 40% of the datacenter cost. This ratio increases with tiered energy pricing from $< 1\%$ to 25%, since this pricing scheme reduces datacenter power consumption and magnify the network cost.

For this policy, we also calculate the profit of network providers based on the energy costs associated with the WAN. Table 6 shows the profit normalized against

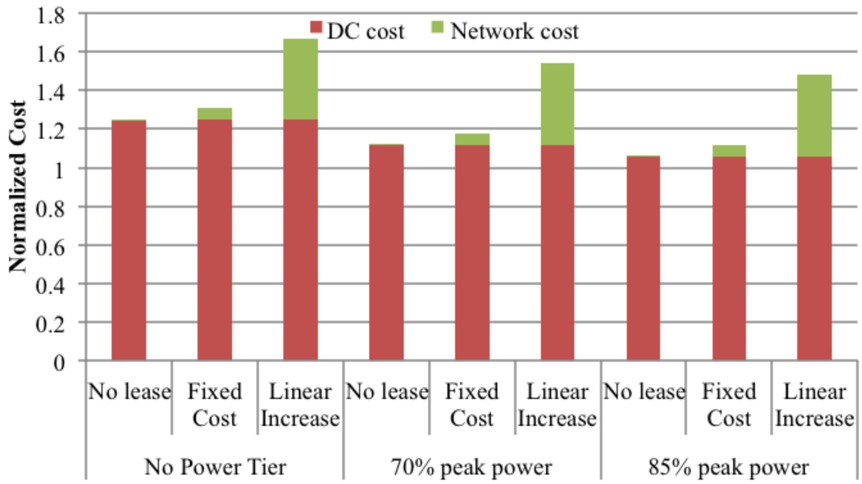

Fig. 7 Normalized performance maximization algorithm costs for datacenters and network

fixed bandwidth cost and non-energy proportional routers. Energy proportionality of routers enables up to 37% more profit for network providers with ideal power curves and 20% with step proportionality WAN router power curve. We also observe that different power tier levels do not affect the savings of the network provider because the migration is based only on green energy availability in other locations.

Table 6 Profit of network providers for performance maximization with different router energy proportionality schemes

Network Cost Function	Profit			
	Non-prop	*Step*	*Smooth*	*Ideal*
Fixed Cost	1x	1.2x	1.2x	1.4x
Linear Increase	4.5x	6.7x	6.8x	6.9x

6.3 Cost minimization using migration

The main goal of the cost minimization policy is to maximize green energy usage and then leverage as much as possible inexpensive brown energy. Also, we show the impact of energy proportional servers to quantify the policy's benefit in future systems.

Unlike *performance maximization*, *cost minimization* does not transfer extra jobs, and thus, does not obtain any performance improvement. Furthermore, the overhead of network transfer decreases the performance of MapReduce jobs. We observe 23.8 min average job completion time for MapReduce jobs, 4.5% worse than the no migration case with green energy efficiency of 66%, a 7% improvement over no migration, with no performance overhead for service jobs.

In Figure 8, we show the impact of energy proportionality and tiered energy pricing to our model, normalized against the no migration case. We observe a 10% decrease in total cost when tiered energy pricing is incorporated into the model. Cost reduction grows to 15% when energy proportional servers are used. This shows the potential of cost minimization method in the future when servers become more energy proportional.

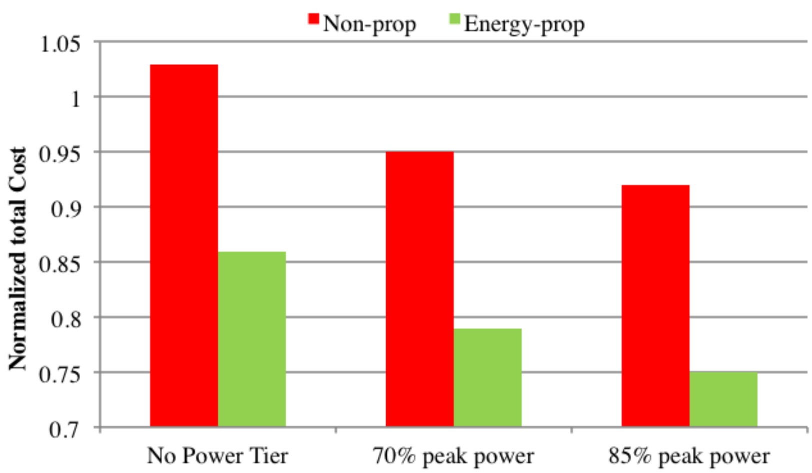

Fig. 8 Normalized cost minimization algorithm costs with different power tier levels and energy proportionality

We also analyze how the total cost of datacenters changes if the network is leased. Unlike the *performance maximization* policy, we prevent migration if the cost is higher than the potential savings. Figure 9 shows the results of this analysis, and additionally incorporates server energy proportionality. We use the same coefficients for the network cost functions as in the previous case. Neglecting the cost of network leasing can result in up to 15% error. The network costs are up to 17% of the datacenter cost, which is significantly less than results we saw with the performance maximization, where it is up to 40%. This is mainly because this policy sacrifices a potential increase in performance if the cost of a data transfer outweighs the cost savings. Figure 9 also shows how bandwidth utilization changes with different power tier levels and network lease options. First, as network costs become more dominant, bandwidth utilization decreases due to a growth in unfeasible data transfers. As a result, if the lease cost is not modeled, the average band-width utilization has up to 60% error. Introducing tiered power levels decreases network utilization because they create a more balanced energy cost scheme across datacenters. Table 7 shows the normalized profit of the network providers. The cost minimization policy inherently limits network profits, since it only allows financially profitable transfers.

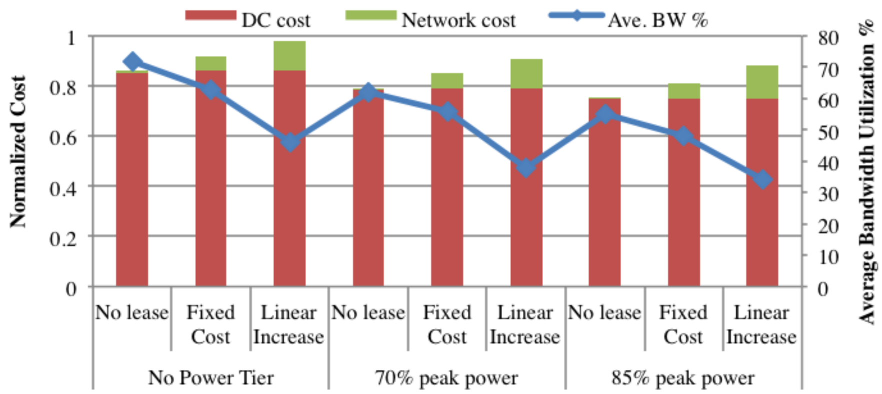

Fig. 9 Normalized total cost and utilization for cost min. with different power tier levels and network lease options using energy proportional servers

Table 7 Profit of network providers for cost. min. with different router energy prop. and with server energy prop.

Network Cost Function	Profit							
	Non-prop		*Step*		*Smooth*		*Ideal*	
	85%	*70%*	*85%*	*70%*	*85%*	*70%*	*85%*	*70%*
Fixed Cost	1x	1.2x	1.2x	1.4x	1.2x	1.4x	1.4x	1.6x
Linear Increase	2.2x	2.45x	3.26x	3.6x	3.4x	3.8x	3.5x	3.9x

6.4 Cost minimization using a green energy aware network

We now investigate the cost minimization policy incorporating green energy aware routing (GEAR). Instead of simply selecting the shortest path between two datacenters, GEAR chooses the path with the least brown energy need. As we only change the network routing policy for this scenario, datacenter cost values are similar compared to the previous case. An energy-aware network provides several benefits. Reducing brown energy costs of the WAN improves overall networking costs for both providers and datacenters. It also provides a viable alternative for datacenters, opting for cheaper green energy at the cost of GEAR's slightly increased network latency. Also, as network elements become more energy proportional in the future, we expect the energy savings obtained by GEAR to be more prominent.

Figure 10 compares SPR and GEAR in terms of router energy consumption and network provider profit, using fixed cost per bandwidth. GEAR with energy proportionality increases profits by 50% compared to the base case (non-proportional, SPR), and provides profit for all proportionality schemes. Without energy proportional routers, GEARs brown energy consumption is slightly lower than SPR (62% vs. 65% of SPR) with a 3% increase in network delay as a result of occasionally choosing a longer path, though with negligible overall effect on the job completion time.

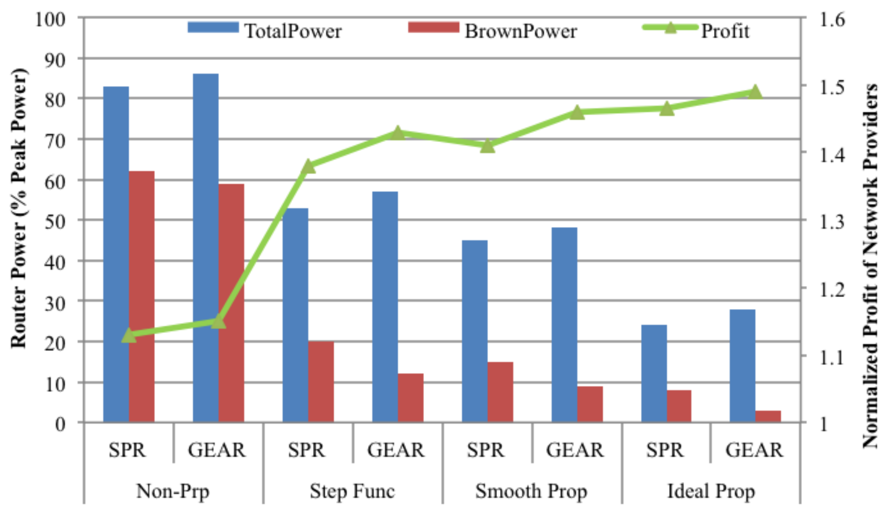

Fig. 10 Comparison between SPR and GEAR energy consumption of routers and network profit with different energy proportionality schemes

7 Discussion

In this section, we first recap the most important results of the above case studies. We then compare our methodology with previous work, and explore the lessons learned with our analysis. Table 8 shows the comparison among the methods discussed in the previous sections. Our performance maximization algorithm uniquely leverages both workload and green energy differences across distributed datacenters to maximize both throughput (27% improvement) and green energy efficiency (44% increase). We also demonstrate that the same variations in workloads and green energy can be leveraged for cost minimization, where our algorithm utilizes tiered energy pricing, and both migration and green energy aware routing. The results show up to 19% reduction in energy cost and 7% improvement in green energy usage while meeting QoS of latency sensitive applications, and increasing job completion time of batch jobs by only 4%. Additionally, the comprehensive and novel aspects of our model provide a level of realistic simulation that previous models do not exhibit to make a complete analysis.

Table 8 Comparison of different policies with respect to total cost, MapReduce performance and green energy usage

		Non-Energy Proportional Servers			**Energy Proportional Servers**		
Policy	**MapReduce Job Completion Time**	**Power Tier**	**Total Energy Cost**	**Green Energy Usage**	**Power Tier**	**Total Energy Cost**	**Green Energy Usage**
No Mig.	**22.8 min**	-	**1.22x**	**59%**	-	**0.99x**	**47%**
		85%	**1x**		85%	**0.85x**	
		70%	**1.10x**		70%	**0.92x**	
Perf. Max.	**16.8 min**	-	**1.25x**	**85%**	-	**1.03x**	**80%**
		85%	**1.06x**		85%	**1x**	
		70%	**1.12x**		70%	**1.05x**	
Cost Min.	**23.8 min**	-	**1.03x**	**66%**	-	**0.86x**	**60%**
		85%	**0.92x**		85%	**0.75x**	
		70%	**0.95x**		70%	**0.79x**	

Green Energy Prediction and Workload Migration Green energy prediction mitigates the inefficiency caused by the variability of renewable sources. We further improve inefficiency by matching our prediction horizon to the long-running batch jobs. The result is better decision making, and as the results indicate, up to 26% improvement in green energy efficiency. Previous work [39], [27] only uses green energy as a method to reduce carbon footprint, and deploy workload migration to improve performance considering load balancing and resource availability [51]. In contrast, we show green energy can also be used to improve performance. We initially propose the idea in [6] for a single datacenter, but now leverage prediction and availability across a network to run extra batch jobs in remote locations. We obtain 27% better batch job completion time compared to no migration with only a 6%-12% increase in total energy cost. Our work is the first to demonstrate the potential of green energy not only as a resource for environmental concerns, but also a means of performance improvement. While cost minimization precludes all potential migrations due to network costs, it still has 7% improvement in green energy usage.

WAN Ownership and Leasing Related work assumes that WAN is part of the datacenter network, or applies static bandwidth costs. However, the WAN may be leased or owned, typically with bandwidth-dependent pricing. Our work is the first to accurately consider such costs. Our first observation is that higher network cost reduces the bandwidth utilization. Secondly, despite increasing network costs with larger cost functions, datacenters can obtain 2-19% cost savings by checking the financial feasibility of each potential migration. In contrast, when the datacenter owns the network, disregarding the initial WAN cost, it achieves up to 22% cost savings.

Tiered Energy Pricing Previous work on minimizing total energy cost, [40], [33], [49] uses grid pricing as either fixed or variable with load. Others [19] attempt to limit datacenter peak load but do not consider how different power levels can affect overall energy cost. Not modeling different cost regions for data center energy consumption may not be correct due to large power consumption of the datacenters. We demonstrate that proposed improvements might be overestimated by up to 20% when accurate pricing is taken into account. Both of our algorithms inherently attempt to remain below tiered power levels in order to avoid higher energy prices, and only exceed those limits when inevitable, i.e. when all datacenters are over-provisioned. Consequently, while our algorithms' performance and cost benefits are tempered by the incorporation of tiered energy pricing, we can still show up to 15% cost savings.

Energy-Proportional Routing We investigate the future of datacenter communication, analyzing the impact of energy proportionality of routers on network provider profit, which has not been explored before. We show that dynamic, green energy aware routing (GEAR) policies can improve energy efficiency by reducing brown energy consumption up to 65%. We quantify that energy proportionality can increase the profit of network providers up to 35% and 57% with fixed and linear policies, respectively. The difference in profit between an implementable proportionality scheme (i.e. step-function) and the ideal case is between 5-17% and decreases with increasing network lease costs. The key observation is that router energy-proportionality schemes can increase profits significantly if deployed, and that GEAR can decrease network brown energy use up to 3x with energy proportionality [2] with negligible performance impact.

Power-Proportional Computing for Future Systems Current datacenter hardware is highly non-energy proportional, resulting in power-inefficient systems. There has been recent work [35] on designing energy-proportional elements. Our work quantifies the benefits of this trend in both major aspects of a datacenter network: servers and network elements. We show the benefit of optimizing the components individually and together into an ideal energy-proportional system, with up to 30% energy savings despite being limited by tiered energy pricing and network contracts. Table 8 quantifies both the impact of such systems, and the continued benefit of our algorithms in a power-proportional environment.

8 Conclusion

Energy efficiency and green energy usage in datacenters and their networks has gained importance as their energy consumption, carbon emissions, and costs have increased dramatically. Previous work leverages geographically separated datacenters by migrating workloads over WAN, leveraging demand and price differences. However, the work neglects several key cost and energy contributions: the financial network, and consequently, data migration costs, focusing solely on latency and QoS costs. Additionally, these publications assume a simpler, and ultimately inaccurate, model for datacenter energy costs. To counteract this, we explore tiered energy pricing for datacenters, network cost models and the costs of owning/leasing a datacenter WAN. We then quantify the inaccuracy of conclusions (25-40% error) drawn when these two are omitted. We solve the variability problem of green energy by using novel prediction algorithms, and subsequently develop algorithms for energy management, focusing on 1) performance maximization, and 2) cost minimization. With the performance maximization algorithm, we demonstrate the ability to leverage green energy to actually improve workload throughput, rather than simply reducing the operational costs. We explore and quantify up to 22% cost savings when realistic WAN costs are incorporated, and up to 65% reduction in network costs when deploying Green Energy Aware Routing (GEAR). We further explore the viability of our new algorithms in the face of emerging technologies in datacenter infrastructure, showing continued benefit of both the performance maximization and the cost minimization algorithms in the presence of energy proportional computing and communication. Our results show that our performance maximization improves batch job performance by 27% while meeting the QoS of services, and that our cost minimization policy decreases overall energy cost by 16%, even when tempered by realistic energy pricing schemes and networking contracts. In future work, we will look to merge the two algorithms to create a balance between the performance gains and cost reduction, for an optimal cost-performance for different datacenter configurations.

9 Acknowledgment

This work was sponsored in part by the Multiscale Systems Center (MuSyC), National Science Foundation (NSF) Project GreenLight, Energy Sciences Network (ESnet), NSF ERC CIAN, NSF Variability Expedition, NSF Flash Gordon, CNS, NSF IRNC TransLight/StarLight, Oracle, Microsoft and Google. We also want to thank Chin Guok for his guidance on path computation algorithms.

References

1. Ajay Mahimkar, Angela Chiu, Robert Doverspike, Mark D. Feuer, Peter Magill, Emmanuil Mavrogiorgis, Jorge Pastor, Sheryl L. Woodward, and Jennifer Yates. 2011. Bandwidth on demand for inter-data center communication. In Proceedings of the 10th ACM Workshop on Hot Topics in Networks (HotNets-X). ACM, New York, NY, USA, Article 24. DOI=10.1145/2070562.2070586
2. Aksanli, B.; Rosing, T.S.; Monga, I, "Benefits of green energy and proportionality in high speed wide area networks connecting data centers," Design, Automation & Test in Europe Conference & Exhibition (DATE), 2012 , vol., no., pp.175,180, 12-16 March 2012 doi: 10.1109/DATE.2012.6176458

3. Ananth Narayan Sankaranarayanan, Somsubhra Sharangi, and Alexandra Fedorova. 2011. Global cost diversity aware dispatch algorithm for heterogeneous data centers. In Proceedings of the 2nd ACM/SPEC International Conference on Performance engineering (ICPE '11). ACM, New York, NY, USA, 289-294. DOI=10.1145/1958746.1958787
4. Asfandyar Qureshi, Rick Weber, Hari Balakrishnan, John Guttag, and Bruce Maggs. 2009. Cutting the electric bill for internet-scale systems. In Proceedings of the ACM SIGCOMM 2009 conference on Data communication (SIGCOMM '09). ACM, New York, NY, USA, 123-134. DOI=10.1145/1592568.1592584
5. AT&T. 2007. http://www.att.com/gen/pressroom?pid=4800&cdvn=news&newsarticleid=24555.
6. Baris Aksanli, Jagannathan Venkatesh, Liuyi Zhang, and Tajana Rosing. 2011. Utilizing green energy prediction to schedule mixed batch and service jobs in data centers. In Proceedings of the 4th Workshop on Power-Aware Computing and Systems (HotPower '11). ACM, New York, NY, USA, , Article 5. DOI=10.1145/2039252.2039257
7. Barroso, L. A., & Hlzle, U. (2009). The datacenter as a computer: An introduction to the design of warehouse-scale machines. Synthesis lectures on computer architecture, 4(1), 1-108.
8. Bergler, B.; Preschern, C.; Reiter, A; Kraxberger, S., "Cost-Effective Routing for a Greener Internet," Green Computing and Communications (GreenCom), 2010 IEEE/ACM Int'l Conference on & Int'l Conference on Cyber, Physical and Social Computing (CPSCom) , vol., no., pp.276,283, 18-20 Dec. 2010 doi: 10.1109/GreenCom-CPSCom.2010.112
9. California ISO. (2012). Retrieved from OASIS: http://oasis.caiso.com/
10. Chow, C. W., Urquhart, B., Lave, M., Dominguez, A., Kleissl, J., Shields, J., & Washom, B. 2011. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. Solar Energy, 85(11), 2881-2893.
11. Dennis Abts, Michael R. Marty, Philip M. Wells, Peter Klausler, and Hong Liu. 2010. Energy proportional datacenter networks. In Proceedings of the 37th annual international symposium on Computer architecture (ISCA '10). ACM, New York, NY, USA, 338-347. DOI=10.1145/1815961.1816004
12. Dondi, D., Zappi, P., & Rosing, T. (2010). A Scheduling Algorithm for High-Performance Monitoring WSN with Hybrid Energy Harvester. ISLPED 2010.
13. Economou, D., Rivoire, S., Kozyrakis, C., & Ranganathan, P. (2006). Full-system power analysis and modeling for server environments. International Symposium on Computer Architecture-IEEE.
14. Franco Travostino, Paul Daspit, Leon Gommans, Chetan Jog, Cees de Laat, Joe Mambretti, Inder Monga, Bas van Oudenaarde, Satish Raghunath, and Phil Yonghui Wang. 2006. Seamless live migration of virtual machines over the MAN/WAN. Future Gener. Comput. Syst. 22, 8 (October 2006), 901-907. DOI=10.1016/j.future.2006.03.007
15. Fratto, M. 2009. http://www.networkcomputing.com/data-center/229503323
16. Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M., & Draxl, C. 2011. The State-Of-The-Art in Short-Term Prediction of Wind Power: A Literature Overview, 2nd edition. ANEMOS.plus.
17. Global Action Plan Report, An inefficient truth. 2007. http://www.globalactionplan.org.uk/
18. Gaurav Dhiman, Giacomo Marchetti, and Tajana Rosing. 2009. vGreen: a system for energy efficient computing in virtualized environments. In Proceedings of the 2009 ACM/IEEE international symposium on Low power electronics and design (ISLPED '09). ACM, New York, NY, USA, 243-248. DOI=10.1145/1594233.1594292
19. Gmach, D.; Rolia, J.; Bash, C.; Yuan Chen; Christian, T.; Shah, A; Sharma, R.; Zhikui Wang, "Capacity planning and power management to exploit sustainable energy," Network and Service Management (CNSM), 2010 International Conference on , vol., no., pp.96,103, 25-29 Oct. 2010 doi: 10.1109/CNSM.2010.5691329
20. Google. (n.d.). Retrieved from: http://www.google.com/transparencyreport/traffic/
21. Guok, C.P.; Robertson, D.W.; Chaniotakisy, E.; Thompson, M.R.; Johnston, W.; Tierney, B., "A User Driven Dynamic Circuit Network Implementation," GLOBECOM Workshops, 2008 IEEE , vol., no., pp.1,5, Nov. 30 2008-Dec. 4 2008 doi: 10.1109/GLOCOMW.2008.ECP.14

22. Hadoop. (n.d.). Retrieved from: http://hadoop.apache.org/
23. Holt, C. C. 2004. Forecasting seasonals and trends by exponentially weighted moving averages. International Journal of Forecasting, 20(1), 5-10.
24. Intel. (n.d.). Retrieved from: www.intel.com/Xeon
25. Jonathan Koomey. 2011. Growth in Data center electricity use 2005 to 2010. Oakland, CA: Analytics Press. August 1. http://www.analyticspress.com/datacenters.html
26. Kien Le; Bianchini, R.; Nguyen, T.D.; Bilgir, O.; Martonosi, M., "Capping the brown energy consumption of Internet services at low cost," Green Computing Conference, 2010 International , vol., no., pp.3,14, 15-18 Aug. 2010 doi: 10.1109/GREENCOMP.2010.5598305
27. Kien Le, Ozlem Bilgir, Ricardo Bianchini, Margaret Martonosi, and Thu D. Nguyen. 2010. Managing the cost, energy consumption, and carbon footprint of internet services. In Proceedings of the ACM SIGMETRICS international conference on Measurement and modeling of computer systems (SIGMETRICS '10). ACM, New York, NY, USA, 357-358. DOI=10.1145/1811039.1811085
28. Kontorinis, V.; Zhang, L.E.; Aksanli, B.; Sampson, J.; Homayoun, H.; Pettis, E.; Tullsen, D.M.; Simunic Rosing, T., "Managing distributed UPS energy for effective power capping in data centers," Computer Architecture (ISCA), 2012 39th Annual International Symposium on , vol., no., pp.488,499, 9-13 June 2012 doi: 10.1109/ISCA.2012.6237042
29. Krioukov, A., Goebel, C., Alspaugh, S., Chen, Y., Culler, D., and Katz, R. 2011. Integrating Renewable Energy Using Data Analytics Systems: Challenges and Opportunities.
30. Kusiak, A., Zheng, H., & Song, Z. (2009). Wind farm power prediction: a datamining approach. Wind Energy, 12(3), 275-293.
31. Kusiak, A; Haiyang Zheng; Zhe Song, "Short-Term Prediction of Wind Farm Power: A Data Mining Approach," Energy Conversion, IEEE Transactions on , vol.24, no.1, pp.125,136, March 2009 doi: 10.1109/TEC.2008.2006552
32. Le, K., Bianchini, R., Martonosi, M., and Nguyen, T. D. 2009. Cost-and energy-aware load distribution across data centers. Proceedings of HotPower, 1-5.
33. Lei Rao; Xue Liu; Le Xie; Wenyu Liu, "Minimizing Electricity Cost: Optimization of Distributed Internet Data Centers in a Multi-Electricity-Market Environment," INFOCOM, 2010 Proceedings IEEE , vol., no., pp.1,9, 14-19 March 2010 doi: 10.1109/INFCOM.2010.5461933
34. Lei Rao, Xue Liu, Marija Ilic, and Jie Liu. 2010. MEC-IDC: joint load balancing and power control for distributed Internet Data Centers. In Proceedings of the 1st ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS '10). ACM, New York, NY, USA, 188-197. DOI=10.1145/1795194.1795220
35. Luiz Andre Barroso, Urs Holzle, "The Case for Energy-Proportional Computing," Computer, vol. 40, no. 12, pp. 33-37, December, 2007
36. Mahendra Kutare, Greg Eisenhauer, Chengwei Wang, Karsten Schwan, Vanish Talwar, and Matthew Wolf. 2010. Monalytics: online monitoring and analytics for managing large scale data centers. In Proceedings of the 7th international conference on Autonomic computing (ICAC '10). ACM, New York, NY, USA, 141-150. DOI=10.1145/1809049.1809073
37. Mankoff, J.; Kravets, R.; Blevis, E., "Some Computer Science Issues in Creating a Sustainable World," Computer , vol.41, no.8, pp.102,105, Aug. 2008 doi: 10.1109/MC.2008.307
38. Miller, R. 2011. Green Data Centers. Data Center Knowledge. http://www.datacenterknowledge.com/archives/category/infrastructure/green-data-centers/
39. Mohsenian-Rad, A. H., & Leon-Garcia, A. (2010). Energy-information transmission tradeoff in green cloud computing. Carbon, 100, 200.
40. Niv Buchbinder, Navendu Jain, and Ishai Menache. 2011. Online job-migration for reducing the electricity bill in the cloud. In Proceedings of the 10th international IFIP TC 6 conference on Networking - Volume Part I (NETWORKING'11), Jordi Domingo-Pascual, Pietro Manzoni, Ana Pont, Sergio Palazzo, and Caterina Scoglio (Eds.), Vol. Part I. Springer-Verlag, Berlin, Heidelberg, 172-185.
41. NREL Solar Maps. (2012). Retrieved from NREL: Available: http://www.nrel.gov/gis/solar.html

42. NREL Wind Maps. (2012). Retrieved from NREL: Available: http://www.nrel.gov/gis/wind.html
43. Piorno, J.R.; Bergonzini, C.; Atienza, D.; Rosing, T.S., "Prediction and management in energy harvested wireless sensor nodes," Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, 2009. Wireless VITAE 2009. 1st International Conference on , vol., no., pp.6,10, 17-20 May 2009 doi: 10.1109/WIRELESSVITAE.2009.5172412
44. Priya Mahadevan, Puneet Sharma, Sujata Banerjee, and Parthasarathy Ranganathan. 2009. A Power Benchmarking Framework for Network Devices. In Proceedings of the 8th International IFIP-TC 6 Networking Conference (NETWORKING '09), Luigi Fratta, Henning Schulzrinne, Yutaka Takahashi, and Otto Spaniol (Eds.). Springer-Verlag, Berlin, Heidelberg, 795-808.
45. RUBiS. (n.d.). Retrieved from http://rubis.ow2.org/
46. Sanchez, I. 2006. Short-term prediction of wind energy production. International Journal of Forecasting, 22(1), 43-56.
47. Tucker, RodneyS.; Baliga, J.; Ayre, R.W.A; Hinton, K.; Sorin, W.V., "Energy consumption in IP networks," Optical Communication, 2008. ECOC 2008. 34th European Conference on , vol., no., pp.1,1, 21-25 Sept. 2008 doi: 10.1109/ECOC.2008.4729202
48. U.S. Energy Information Administration. Electric Power Monthly. (n.d.). Retrieved from http://www.eia.gov/electricity/monthly/
49. Yanpei Chen; Ganapathi, A; Griffith, R.; Katz, R., "The Case for Evaluating MapReduce Performance Using Workload Suites," Modeling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS), 2011 IEEE 19th International Symposium on , vol., no., pp.390,399, 25-27 July 2011 doi: 10.1109/MASCOTS.2011.12
50. Yanwei Zhang, Yefu Wang, and Xiaorui Wang. 2011. Capping the electricity cost of cloud-scale data centers with impacts on power markets. In Proceedings of the 20th international symposium on High performance distributed computing (HPDC '11). ACM, New York, NY, USA, 271-272. DOI=10.1145/1996130.1996170
51. Zhenhua Liu, Minghong Lin, Adam Wierman, Steven H. Low, and Lachlan L.H. Andrew. 2011. Greening geographical load balancing. In Proceedings of the ACM SIGMETRICS joint international conference on Measurement and modeling of computer systems (SIGMETRICS '11). ACM, New York, NY, USA, 233-244. DOI=10.1145/1993744.1993767