# Energy and Environmental Aspects of Data Centers

Sabrina Spatari[1,*], Nagarajan Kandasamy[2], Dara Kusic[3], Eugenia V. Ellis[4], Jin Wen[1]

[1]*Dept. of Civil, Architectural, and Environmental Engineering, Drexel University, Philadelphia, USA*
[2]*Dept. of Electrical and Computer Engineering, Drexel University, Philadelphia, USA*
[3]*Coriell Institute, Camden, USA*
[4]*Dept. of Architecture & Interiors, Drexel University, Philadelphia, USA*
** Corresponding author. Tel: +215-571-3557, Fax: +215 895-1363, E-mail: spatari@drexel.edu*

**Abstract:** Data centers have become an essential operational component of nearly every sector of the economy, and as a result growing consumers of energy and emitters of greenhouse gases (GHGs). Developing strategies for optimizing power usage and reducing the associated life cycle GHG emissions are critical priorities for meeting climate policy objectives. We investigate data center power management through virtualization, a technique that consolidates data center workloads onto fewer computing resources within a data center and deploys computing resources only as needed. Based on an experimentally validated dynamic resource provisioning framework applied to a small scale computing cluster at Drexel University that employed lookahead control, a control scheme using virtualization demonstrated a 25% reduction in power consumption over a 24-hour period. Using the power savings results from the virtualization experiments, and extrapolating those savings to a medium-sized data center that hosts 500 servers, we estimate the avoided life cycle GHG emissions for implementing a virtualization strategy in hourly time-steps for marginal and average electricity units over a 24-hour day during the month of August, when electricity loads are typically highest for the year. Results from this work show virtualization could avoid the emission of approximately 0.8 to 1.2 metric tons $CO_2$e/day.

***Keywords:*** *Energy, Power management, Buildings, Information technology, Life cycle assessment*

## 1. Introduction

Data centers have become an essential operational component of nearly every sector of the economy, and are additionally growing consumers of energy, and as a result, a burgeoning source of greenhouse gas (GHG) emissions. In 2008, data centers worldwide emitted 170 million tons of $CO_2$, an output on parallel with the total GHG inventory of countries such as Argentina and the Netherlands. Moreover, data center GHG emissions are expected to grow four-fold by 2020 and surpass those of the airline industry [1]. Therefore, developing strategies for optimizing power usage and reducing the associated life cycle GHG emissions are critical priorities for meeting climate policy objectives.

This paper investigates use of new techniques for server power management and validates them using a small-scale computing testbed. The validation experiments are integrated with environmental life cycle models that evaluate the consequential reduction in life cycle GHG emissions of computing equipment and data centers as a whole as a result of the power management strategy in combination with expected input sources of electricity and regional electricity mixes. We accomplish this by coupling the data center power optimization strategy with a life cycle model of electricity supply that examines the average electricity mix and marginal units of power supply over a 24-hour period. The power management strategy we examine is known as virtualization, a technique that consolidates multiple online services onto fewer computing resources within a data center and deploys computing resources only as needed.

### *1.1. Background*

Energy management in data centers involves three main components of computer hardware, building, electricity infrastructure: power load distribution of the data operations; cooling systems employed to control ambient conditions in the buildings that house the computers (the HVAC systems); and the electric power supply system (the electricity grid) and sources that make up each composite unit of electricity (measured in kWh) delivered to the data center, consisting of hydro-electric, nuclear, coal, natural gas, oil, and renewable sources (e.g., wind power, biomass, geothermal, etc). In addition to the in-use consumption of energy and emission of GHGs, the three components of a data center (building, hardware, electric utility infrastructure) carry "upstream" energy and GHG emissions owing to the processes and resource inputs and capital equipment used to produce (mine, manufacture, transport, and construct) a data center.

A typical data center serves a variety of companies and users, and the computing resources needed to support such a wide range of online services leaves server rooms in a state of "sprawl" with under-utilized resources. Moreover, each new service to be supported often results in the acquisition of new hardware, leading to server utilization levels, by some estimates, at less than 20%. With energy costs rising and society's need to reduce energy consumption, it is imprudent to continue server sprawl at its current pace.

Virtualization provides a promising approach to consolidating multiple online services onto fewer computing resources within a data center. This technology allows a single server to be shared among multiple performance-isolated platforms known as virtual machines (VMs), where each virtual machine can, in turn, host multiple enterprise applications. Virtualization also enables on-demand or utility computing, a dynamic resource provisioning model in which computing resources such as the central processing unit (CPU) and memory are made available to applications only as needed and not allocated statically based on the peak workload. By dynamically provisioning virtual machines, consolidating the workload, and turning servers on and off as needed, data center operators can maintain service-level agreements (SLAs) with clients while achieving higher server utilization and energy efficiency.

In this research paper, we apply an experimentally validated dynamic resource provisioning framework for integrated power and performance management in virtualized computing environments developed by Kusic and Kandasamy [2,3,4]. Prior research by the authors demonstrated the novelty of the application of advanced control, optimization, and mathematical programming concepts to provide the necessary theoretical basis for this framework. The authors posed the power/performance management problem as one of sequential optimization under uncertainty and solve this problem using limited lookahead control (LLC), an adaptation of the well-known model-predictive control approach [2]. The framework solves an online optimization problem that maximizes the performance objective over a given prediction horizon, and then periodically rolls this horizon forward. The LLC approach allows for multiple objectives (such as power and performance) to be represented as optimization problems under explicit operating constraints and solved for every control step. It is also applicable to computing systems with complex non-linear behavior where tuning options must be chosen from a finite set at any given time (such as the number of physical machines and/or VMs to power up/down). Experimental results obtained using a small-scale cluster show significant promise that LLC can systematically address performance/power problems within the highly dynamic operating environment of a data center. Table 1 summarizes results from prior research that demonstrated that the server cluster, which hosted six heterogeneous servers that host multiple online services, when managed using LLC saved, on average, 26% in power consumption costs over a 24-hour period, when compared to the uncontrolled case when no servers are ever switched off.

*Table 1 Control performance, measured as average energy savings and SLA violations, for different workloads[a]*

| Workload | Total Energy Savings (%) | % SLA Violations (Silver) | % SLA Violations (Gold) |
|---|---|---|---|
| Workload 1 | 18 | 3.2 | 2.3 |
| Workload 2 | 17 | 1.2 | 0.5 |
| Workload 3 | 17 | 1.4 | 0.4 |
| Workload 4 | 45 | 1.1 | 0.2 |
| Workload 5 | 32 | 3.5 | 1.8 |

[a] Notes: The cluster hosts two services, termed “Gold” and “Silver” enabled by the Trade6 and DVDStore applications. The services generate revenue as per a non-linear pricing scheme that relates the achieved response time-300 ms for each Gold request and 200 ms for each Silver request-to a dollar value. Response times below the threshold result in a reward paid to the service provider, while response times violating the SLA result in the provider paying a penalty to the client.

## 2. Methodology

We apply life cycle assessment (LCA) methods following ISO 2006 [5] procedures to examine the potential for reducing life cycle greenhouse gas (GHG) emissions when employing power management via virtualization. We examine avoided electricity consumption and GHG emissions based on hourly marginal units of electricity in the Pennsylvania-New Jersey-Maryland (PJM) power mix, and in this way use a consequential LCA approach to the problem [6]. The system boundary investigated consists of the data center work load management on 500 central processing units (CPU), overhead building HVAC needs, and the power grid mix that comprises power supply to the data center (Figure 1). A complete LCA model of the system performance outlined in Figure 1 would normally account for the “upstream” energy and associated GHG emissions of the building, computer hardware, and electric utility infrastructure) in addition to their contributions during operation. However, in this paper we limit our scope to operation related energy and environmental performance to isolate the performance of the virtualization strategy independent of system attributes, but we evaluate the changes in life cycle performance induced on the system by the virtualization strategy.

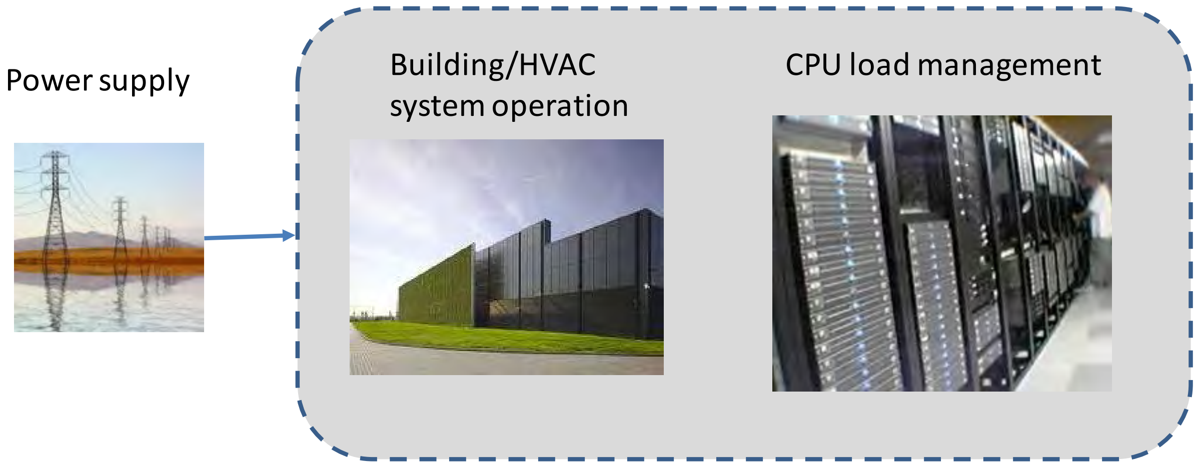


*Figure 1: Systems boundary comprising data center and power supply*

### *2.1. Power Management by Virtualization*

As noted above, experiments conducted on a cluster of Dell PowerEdge servers indicate that the controller achieves a 25% to 46% reduction in power consumption costs for six different workloads over a 24-hour period when compared to an uncontrolled system while achieving the desired QoS goals.

Figure 2 shows the LLC scheme where we obtain the control actions governing system operation by optimizing the forecast system behavior for the performance metric over a limited prediction horizon [3]. The controller obtains the sequence of control actions that results in the best system behavior over this horizon and applies the first action within this sequence as input during the current time instant. It then discards the rest and repeats this process at each time step. Thus, in a predictive-control design, the controller optimizes the performance metric at each sampling-time instance, taking into account future variations in the environment inputs and their effects on system behavior.

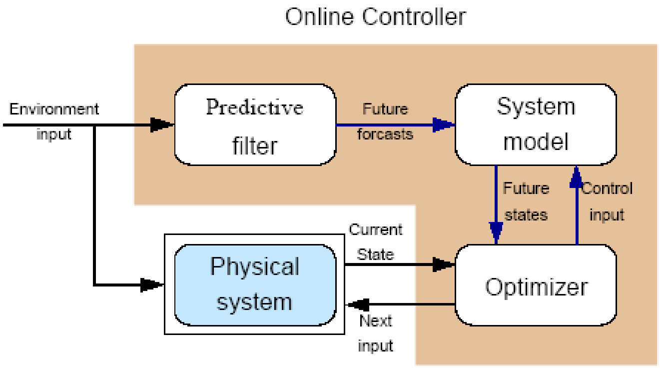


*Fig. 2: Structure of the limited lookahead controller*

The computer testbed cluster, when managed using LLC saves, on average, 26% in power consumption costs over a 24-hour period, when compared to the uncontrolled case when no servers are ever switched off. Also, when the incoming workload is noisy, the risk-aware controller provides superior performance compared to a risk-neutral one by reducing both SLA violations and host switching activity.

### *2.2. Life Cycle Inventory Analysis*

We constructed a life cycle assessment (LCA) model focused on the consequential GHG emissions avoided through implementation of the virtualization experiments scaled to a small-medium sized data center hosting 500 CPUs. We applied tests from workload simulations that achieved a 41% energy savings relative to the uncontrolled case noted in section 2.2 and assume that power savings scale linearly from the testbed cluster experiments to the 500-CPU data center. For the LCA model that we construct here, we consider the reduced demand for power to the data center resulting from consolidating workloads onto fewer machines. Therefore with respect to unit processes considered in the LCA model, we take into account only the power supply needed for the CPUs in the data center, and do not take into account optimizing the HVAC system shown in Fig. 1. However, we note that a control scheme that seeks to minimize overall energy consumption can include inputs from the building cooling systems and environmental ambient conditions.

## 3. Results

A systematic analysis of energy savings allowed through optimization of power management in a data center requires consideration of the electricity supplying the data center and the upstream fuel/resource extraction and transport of energy sources to individual power plants (e.g., coal, nuclear, natural gas, and fuel oil). Typically LCA models employ life cycle inventory (LCI) data that describe electricity grid mixes averaged over time. Over a year, the national U.S. electricity grid tends to average out to approximately 51% coal, 21% nuclear, 16% natural gas, 7% hydro-electric, 3% petroleum, and 3% other, including renewable sources [7]. However, around the different regions of the U.S., regional averages are known and can be used to estimate net life cycle emissions related to power consumption.

LCA attempts to quantify the resource intensity and damages associated directly with the flow of resources and wastes associated with products, processes, and activities, or more generally, "systems". Analysts have tried for some time to track inputs and outputs across different economic sectors in order to capture actual energy and materials consumed for those systems. The structure of electric power distribution within a regional grid does not make it possible to trace individual electrons from source to sink, which is why analysts use average electricity mixes. In the case of data centers, a better approach for understanding power consumption by source, is to track electricity consumption by time of day, as this would show how the data center consumes electricity during peak and off-peak times, and thus the implications for using coal, the highest carbon emitting power sources, versus natural gas, the more efficient of the carbon-emitting sources.

Another debate in LCA literature regarding the trace of electricity usage is the question of whether to count the marginal or the average unit of electricity output. Some argue that each additional unit of demand placed into an electricity grid should correspond to the marginal unit consumed. For example, if a new data center is constructed and it sources its power from the Northeast electricity grid, the marginal unit approach would assume that electricity consumed came from the last kWh of output, the marginal source, since the overall mix at any given time is already meeting existing demand. Analyzing the problem this way, we would expect that a new data center built into a region and relying on electricity supply, would therefore use the marginal unit at each time step during the day. Put another way, any savings in energy from the data center would save GHGs from the marginal unit of output. For the majority of the peak hours, of the day, this could reduce coal-sourced power. At mid-peak hours this may reduce lower-GHG emitting natural gas sources, and in off-peak hours it may save coal-GHG emissions. Approaching the problem this way, we took data from a typical power supply curve on a peak August month, approximated the electricity mix based on peak load distribution curves (Figure 3), and then approximated the mix and GHG savings or avoided coal power based on a data center optimization strategy.

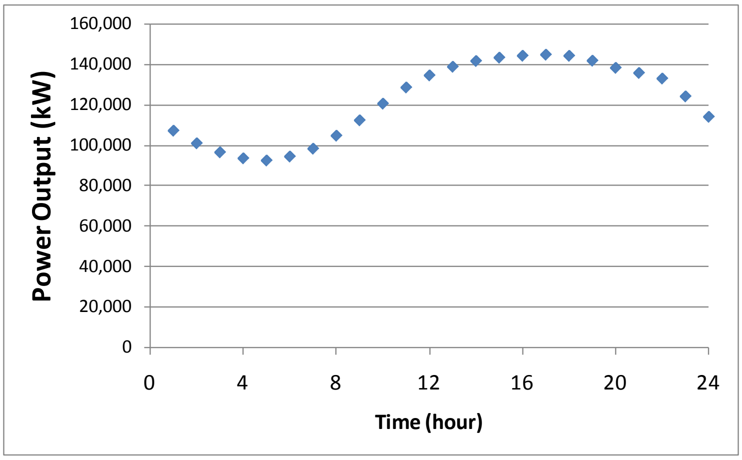


*Figure 3: Peak load distribution in August 2006 (source: PJM)*

The energy savings over the 24 hour day allow up to 61% power savings during late hours and early afternoon, when marginal sources shift between coal and natural gas. Figure 4 shows the energy savings over the 24-hour period resulting from optimizing workloads in the data center (see References [3] and [4] for further detail on the workload trace and optimization scheme applied). The largest energy savings occur during hours 3 to 15 of the day, corresponding largely with peak power demand times. Accordingly, there could be significant savings in net life cycle GHG emissions from fossil energy sources.

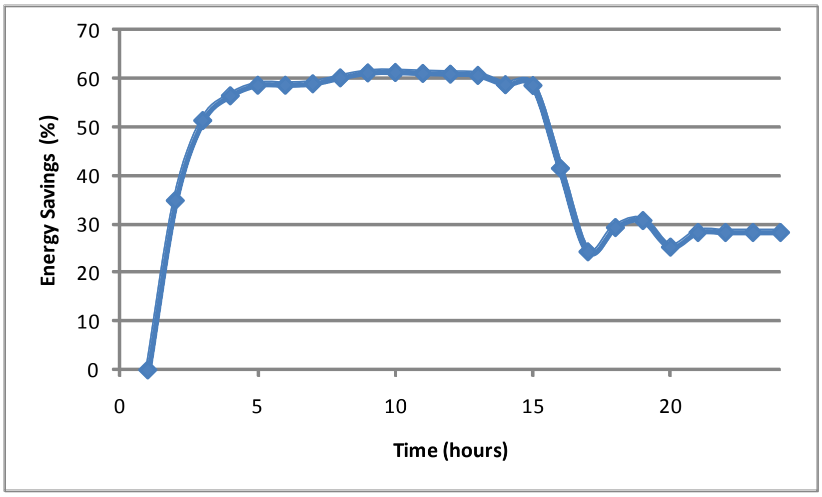


*Figure 4: Energy Savings over a 24-hour period estimated in 2-minute intervals*

We combined the hour-by-hour energy savings estimates with the life cycle electricity supply measured in average and marginal units based on the PJM electricity grid. Figure 5 shows the avoided GHG emissions from deploying the virtualization strategy. We see that counting the marginal unit of electricity results in a larger estimate of avoided GHG emissions. This is especially evident between hours 4 through 11, and is the result of an expected reduction in GHG emissions from coal-based electricity, the marginal unit used during that time interval.

When we sum up the avoided GHG emissions over the day by integrating over the two curves shown in Figure 5, we find that a medium sized data center can avoid 0.8 metric tons $CO_2e$ $day^{-1}$ and 1.2 metric tons $CO_2e$ $day^{-1}$ when counting the average and marginal units of electricity avoided, respectively.

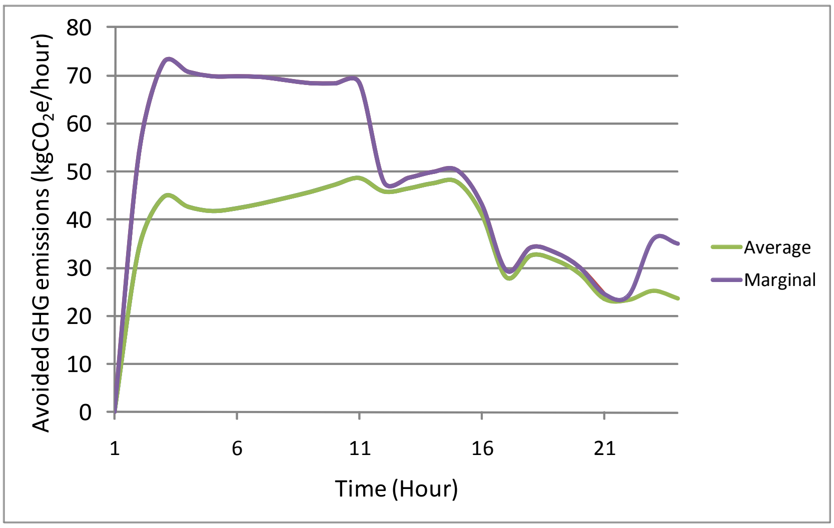


*Figure 5: Estimated avoided greenhouse gas emissions resulting from Energy Savings over a 24-hour period estimated in 2-minute intervals*

## 4. Discussion

This research demonstrates the usefulness of LCA models to estimate the full set of resources saved and environmental damages avoided by power management in data centers. We focused on energy savings enabled through power management. During runtime, the data center is typically managed to reduce resource consumption from the computing, power and cooling infrastructures, and SLAs are used to define the operational requirements of the infrastructure based on the desired performance. The life cycle approach builds in the upstream resources that supply the data center, and time-of-day analysis further delineates specific resources consumed coupled with typical demand patterns of internet resources.

The model tested herein describes performance of one data center located in a particular climate zone during summer peak power demand times, with electricity supply from that zone. This may not be the case in practice since data centers are located throughout the world and different data centers may be deployed as coupled systems when needed in order to optimize performance constraints other than energy/environmental. Knowing this, there may be opportunities to migrate workloads to data centers around the globe to take advantage of latitudes and climates that have the ability to employ passive cooling, and thereby reduce further the need for energy demand on data center HVAC. Design and control of the data centers may also create synergies with renewable power such as wind that tend to go online at night or potentially other day-time renewable energy sources such as photovoltaic for electricity supply. Through scenario analysis that uses LCA, novel building and locating strategies can be identified to further reduce the energy and carbon intensity of this sector. Understanding the interaction between data center location and real time power consumption is critical to optimizing computer cluster usage, since demand during peak hours tends to source marginal sources of power (coal), which tend to be the highest $CO_2$ emitting sources, rather than low-emission alternative sources of energy. These aspects will be investigated in relation to data center design, resource planning, and operation in future work.

There is much opportunity to improve data center performance as part of building design for specific climate zones and in selecting power aware computing hardware. Addressing building design, cooling strategies may include locating the data center on the ground floor or in underground spaces. Certain design approaches for passive cooling of the data center include use of underfloor air distribution systems with natural convection to create zoned ecosystems around equipment, localized air-handling units to redirect warm air from equipment rooms to other building zones, and outside air for cooling. Much additional energy savings can be achieved through coupling building/architectural design components into the power management and control strategy. Temperature sensors, embedded in the physical infrastructure, can be used to guide resource-provisioning decisions that consider the data center as a whole rather than at the level of an individual cluster(s). Workload can be dynamically managed taking into account the impact of provisioning decisions on the overall operating cost of the data center (for example, cooling costs). Automated migration of a virtualized workload from one set of servers located in a hot zone to a cooler zone in the data center (or one that costs less to cool) has potential to significantly reduce cooling costs and overall energy and GHG emissions.

Building on the preliminary results discussed here, our future work will aim to develop a control framework wherein power/performance criteria can be applied to all aspects of data center operation. Such a framework would aim to analyze the three critical components, hardware, building, and electric utility infrastructure (Figure 1) for designing “green” data centers on a life cycle basis, along with operational control and location choice.

### Acknowledgements

This research was supported by the Drexel Engineering Cities Initiative (DECI) at Drexel University. Spatari thanks A.E. Huemmler of the University of Pennsylvania for comments and discussion on this manuscript.

### References

[1] McKinsey & Co., Report: Revolutionizing Data Center Efficiency, 2008.

[2] D. Kusic and N. Kandasamy, "Risk-Aware Limited Lookahead Control for Dynamic Resource Provisioning in Enterprise Computing Systems," Cluster Computing: Special Issue on Autonomic Computing, v. 10, n. 7, Kluwer Academic Publishers, Dec. 2007, pp. 395-408.

[3] Kusic, D., J. O. Kephart, J. E. Hanson, N. Kandasamy, and G. Jiang, “Power and Performance Management of Virtualized Computing Environments via Lookahead Control," Cluster Computing, vol. 12, no. 1, pp. 1-15, Springer Netherlands, March 2009.

[4] D. Kusic, J. Kephart, J. Hanson, N. Kandasamy and G. Jiang, "Power and Performance Management of Virtualized Computing Environments via Lookahead Control," Proc. IEEE Conf. on Autonomic Computing (ICAC'08), Chicago, IL, Jun. 2008, pp 3-12.

[5] ISO, 2006. ISO 14044: Environmental management — Life cycle assessment — Requirements and guidelines. International Standards Organization, Geneva.

[6] Ekvall, T., Weidema, B.P., 2004. System boundaries and input data in consequential life cycle inventory analysis. International Journal of Life Cycle Assessment 9, 161-171.

[7] Kim, S.; Dale, B., “Life Cycle Inventory Information of the United States Electricity System,” (11/17 pp). The International Journal of Life Cycle Assessment 2005, 10, (4), 294-304.