Cutting the Electric Bill for Internet-Scale Systems

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1145/1594977.1592584">http://dx.doi.org/10.1145/1594977.1592584</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Association for Computing Machinery / ACM Special Interest Group on Data Communications</td>
</tr>
<tr>
<td>Version</td>
<td>Author's final manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Tue Nov 20 16:10:18 EST 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/62585">http://hdl.handle.net/1721.1/62585</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution-Noncommercial-Share Alike 3.0</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by-nc-sa/3.0/">http://creativecommons.org/licenses/by-nc-sa/3.0/</a></td>
</tr>
</tbody>
</table>
ABSTRACT

Energy expenses are becoming an increasingly important fraction of data center operating costs. At the same time, the energy expense per unit of computation can vary significantly between two different locations. In this paper, we characterize the variation due to fluctuating electricity prices and argue that existing distributed systems should be able to exploit this variation for significant economic gains. Electricity prices exhibit both temporal and geographic variation, due to regional demand differences, transmission inefficiencies, and generation diversity. Starting with historical electricity prices, for twenty nine locations in the US, and network traffic data collected on Akamai's CDN, we use simulation to quantify the possible economic gains for a realistic workload. Our results imply that existing systems may be able to save millions of dollars a year in electricity costs, by being cognizant of locational computation cost differences.

Categories and Subject Descriptors
C.2.4 [Computer-Communication Networks]: Distributed Systems

General Terms
Economics, Management, Performance

1. INTRODUCTION

With the rise of “Internet-scale” systems and “cloud computing” services, there is an increasing trend toward massive, geographically distributed systems. The largest of these are made up of hundreds of thousands of servers and several data centers. A large data center may require many megawatts of electricity [1], enough to power thousands of homes.

Millions of dollars must be spent annually on the electricity needed to power one such system. Furthermore, these already large systems are increasing in size at a rapid clip, outpacing data center energy efficiency gains [2], and electricity prices are expected to rise.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Figure 1: Estimated annual electricity costs for large companies (servers and infrastructure) @ $60/MWh. These are conservative estimates, meant to be lower bounds. See §2.1 for derivation details. For scale, we have included the actual 2007 consumption and utility bill for the MIT campus, including dormitories and labs.

We observe that most Internet-scale systems cannot ignore their energy costs. A back-of-the-envelope calculation for Google suggests it consumes more than $38M worth of electricity annually (figure 1). A modest 3% reduction would therefore exceed a million dollars every year. We project that even a smaller system like Akamai’s consumes an estimated $10M worth of electricity annually.

The conventional approach to reducing energy costs has been to reduce the amount of energy consumed [3, 4]. New cooling technologies, architectural redesigns, DC power, multi-core servers, virtualization, and energy-aware load balancing algorithms have all been proposed as ways to reduce the power demands of data centers. That work is complementary to ours.

This paper develops and analyzes a new method to reduce the energy costs of running large Internet-scale systems. It relies on two key observations:

1. *Electricity prices vary.* In those parts of the U.S. with wholesale electricity markets, prices vary on an hourly basis and are often not well correlated at different locations. Moreover, these variations are substantial, as much as a factor of 10 from one hour to the next. If, when computational demand is below peak, we can dynamically move demand (i.e., route service requests) to places with lower prices, we can reduce energy costs.

2. *Large distributed systems already incorporate request routing and replication.* We observe that most Internet-scale systems today are geographically distributed, with...
machines at tens or even hundreds of sites around the world. To provide clients good performance and to tolerate faults, these systems implement some form of dynamic request routing to map clients to servers, and often have mechanisms to replicate the data necessary to process requests at multiple sites.

We hypothesize that by exploiting these observations, large systems can save a significant amount of money, using mechanisms for request routing and replication that they already implement. To explore this hypothesis, we develop a simple cost-aware request routing policy that preferentially maps requests to locations where energy is cheaper.

Our main contribution is to identify the relevance of electricity price differentials to large distributed systems and to estimate the cost savings that could result in practice if the scheme were deployed.

**Problem Specification.** Given a large system composed of server clusters spread out geographically, we wish to map client requests to clusters such that the total electricity cost (in dollars, not Joules) of the system is minimized. For simplicity, we assume that the system is fully replicated. Additionally, we optimize for cost every hour, with no knowledge of the future. This rate of change is slow enough to be compatible with existing routing mechanisms, but fast enough to respond to electricity market fluctuations. Finally, we incorporate bandwidth and performance goals as constraints. Existing frameworks already exist to optimize for bandwidth and performance; modeling them as constraints makes it possible to add our process to the end of the existing optimization pipeline.

Note that our analysis is concerned with reducing cost, not energy. Our approach may route client requests to distant locations to take advantage of cheap energy. These longer paths may cause overall energy consumption to rise slightly.

**Energy Elasticity.** The maximum reduction in cost our approach can achieve hinges on the energy elasticity of the clusters. This is the degree to which the energy consumed by a cluster depends on the load placed on it. Ideally, clusters would draw no power in the absence of load. In the worst case, there would be no difference between the peak power and the idle power of a cluster. Present state-of-the-art systems [5, 6] fall somewhere in the middle, with idle power being around 60% of peak. A system with inelastic clusters is forced to always consume energy everywhere, even in regions with high energy prices. Without adequate elasticity, we cannot effectively route the system’s power demand away from high priced areas.

Zero-idle power could be achieved by aggressively consolidating, turning off under-utilized components, and always activating only the minimum number of machines needed to handle the offered load. At present, achieving this without impacting performance is still an open challenge. However, there is an increasing interest in energy-proportional servers [6] and dynamic server provisioning techniques are being explored by both academics and industry [7, 8, 9, 10, 11].

**Results.** To conduct our analysis, we use trace-driven simulation with real-world hourly (and daily) energy prices obtained from a number of data sources. We look at 39 months of hourly electricity prices from 29 US locations. Our request traces come from the Akamai content distribution network (CDN): we obtained 24-days worth of request traffic data (five-minute load) for each server cluster located at a commercial data center in the U.S. We used these data sets to estimate the performance of our simple cost-aware routing scheme under different constraints.

We show that:

- Existing systems can reduce energy costs by at least 2%, without any increase in bandwidth costs or significant reduction in client performance (assuming a Google-like energy elasticity, an Akamai-like server distribution and 95/5 bandwidth constraints). For large companies this can exceed a million dollars a year.

- Savings rapidly increase with energy elasticity: in a fully elastic system, with relaxed bandwidth constraints, we can reduce energy cost by over 30% (around 13% if we impose strict bandwidth constraints), without a significant increase in client-server distances.

- Allowing client-server distances to increase leads to increased savings. If we remove the distance constraint, a dynamic solution has the potential to beat a static solution (i.e., place all servers in cheapest market) by a substantial margin (45% maximum savings versus 35% maximum savings).

Presently, energy cost-aware routing is relevant only to very large companies. However, as we move forward and the energy elasticity of systems increases, not only will this routing technique become more relevant to the largest systems, but much smaller systems will also be able to achieve meaningful savings.

**Paper Organization.** In the next section, we provide some background on server electricity expenditure and sketch the structure of US energy markets. In section 3 we present data about the variation in regional electric prices. Section 4 describes the Akamai data set used in this paper. Section 5 outlines the energy consumption model used in the simulations covered in section 6. Section 7 considers alternative mechanisms for market participation. Section 8 presents some ideas for future work, before we conclude.

2. BACKGROUND

This section first presents evidence that electricity is becoming an increasingly important economic consideration, and then describes the salient features of the wholesale electricity markets in the U.S.

2.1 The Scale of Electricity Expenditures

In absolute terms, servers consume a substantial amount of electricity. In 2006, servers and data centers accounted for an estimated 61 million MWh, 1.5% of US electricity consumption, costing about 4.5 billion dollars [3]. At worst, by replacing everything with state-of-the-art equipment, we may be able to reduce usage in 2011 to half the current level [3].

Most companies operating Internet-scale systems are secretive about their server deployments and power consumption. Figure 1 shows our estimates for several such companies, based on back-of-the-envelope calculations. The

\[ \text{Energy in Wh} \approx n \cdot (P_{idle} + (P_{peak} - P_{idle}) \cdot U + (PUE - 1) \cdot P_{peak}) \cdot 365 \cdot 24, \]

where \( n \) is server count, \( P_{peak} \) is server peak power in Watts, \( P_{idle} \) is idle power, and \( U \) is average server utilization.
server numbers are from public disclosures for eBay [12] and Rackspace (Q1 2009 earnings report). To calculate energy, we have made the following assumptions: average data center power usage effectiveness (PUE) is 2.0 [3] and is calculated based on peak power; average server utilization is around 30% [6, 7]; average peak server power usage is 250 Watts (based on measurements of actual servers at Akamai); and idle servers draw 60-75% of their peak power [5, 8]. Our numbers for Microsoft are based on company statements [13] and energy figures mentioned in a promotional video [14].

To estimate Google’s power consumption, we assumed 500K servers (based on an old, widely circulated number [13]), operating at 140 Watts each [5], a PUE of 1.3 [4] and average utilization around 30% [6]. Such a system would consume more than $6.3 \times 10^8$ MWh, and would incur an annual electricity bill of nearly $38 million (at $60 per MWh wholesale rate). These numbers are consistent with an independent calculation we can make. comScore estimated that Google performed about 1.2B searches/day in August 2007 [15], and Google officially stated recently that each search takes 1 kJ of energy on average (presumably amortized to include indexing and other costs). Thus, search alone works out to $1 \times 10^9$ MWh in 2007. Google’s servers handle GMail, YouTube, and many other applications, so our earlier estimates seem reasonable. Google may well have more than a million servers [1], so an annual electric bill exceeding $80M wouldn’t be surprising.

Akamai’s electricity costs represent indirect costs not seen by the company itself. Like others who rely on co-location facilities, Akamai seldom pays directly for electricity. Power is mostly built into the billing model, with charges based on provisioned capacity rather than consumption. In section 7 we discuss why our ideas are relevant even to those not directly charged per-unit of electricity they use.

### 2.2 Wholesale Electricity Markets

Although market details differ regionally, this section provides a high-level view of deregulated electricity markets, providing a context for the rest of the paper. The discussion is based on markets in the United States.

**Generation.** Electricity is produced by government utilities and independent power producers from a variety of sources. In the United States, coal dominates (nearly 50%), followed by natural gas (~20%), nuclear power (~20%), and hydroelectric generation (6%) [16].

Different regions may have very different power generation profiles. For example, in 2007, hydroelectric sources accounted for 74% of the power generated in Washington state, while in Texas, 86% of the energy was generated using natural gas and coal.

**Transmission.** Producers and consumers are connected to an electric grid, a complex network of transmission and distribution lines. Electricity cannot be stored easily, so supply and demand must continuously be balanced.

In addition to connecting nearby nodes, the grid can be used to transfer electricity between distant locations. The United States is divided into eight reliability regions, with varying degrees of inter-connectivity. Congestion on the grid, transmission line losses (est. 6% [17] in 2006), and boundaries between regions introduce distribution inefficiencies and limit how electricity can flow.

**Market Structure.** In each region, a pseudo-governmental body, a Regional Transmission Organization (RTO), manages the grid (figure 2). An RTO provides a central authority that sets up and directs the flow of electricity between generators and consumers over the grid. RTOs also provide mechanisms to ensure the short-term reliability of the grid.

Additionally, RTOs administer wholesale electricity markets. While bilateral contracts account for the majority of the electricity that flows over the grid, wholesale electricity trading has been growing rapidly, and presently covers about 40% of total electricity.

Wholesale market participants can trade forward contracts for the delivery of electricity at some specified hour. In order to determine prices for these contracts, RTOs such as PJM use an auctioning mechanism: power producers present supply offers (possibly price sensitive), consumers present demand bids (possibly price sensitive); and a coordinating body determines how electricity should flow and sets prices. The market clearing process sets hourly prices for the different locations in the market. The outcomes depend not only on bids and offers, but also account for a number of constraints (grid-connectivity, reliability, etc.).

Each RTO operates multiple parallel wholesale markets. There are two common market types:

- **Day-ahead markets (futures)** provide hourly prices for delivery during the following day. The outcome is based on expected load.

- **Real-time markets (spot)** are balancing markets where prices are calculated every five minutes or so, based on actual conditions, rather than expectations. Typically, this market accounts for a small fraction of total energy transactions (less than 10% of total in NYISO).

Generally speaking, the most expensive active generation resource determines the market clearing price for each hour. The RTO attempts to meet expected demand by activating the set of resources with the lowest operating costs. When demand is low, the base-load power plants, such as coal and nuclear can fulfill it. When demand rises, additional resources, such as natural gas turbines, need to be activated.

Security constraints, line losses and congestion costs also impact price. When transmission system restrictions, such as line capacities, prevent the least expensive energy supplier from serving demand, congestion is said to exist. More

---

4 A measure of data center energy efficiency.
Figure 3: Daily averages of day-ahead peak prices at different hubs [18]. The elevation in 2008 correlates with record high natural gas prices, and does not affect the hydroelectric dominated Northwest. The Northwest consistently experiences dips near April (this seems to be correlated with seasonal rainfall). Correlated with the global economic downturn, recent prices in all four locations exhibit a downward trend.

Figure 4: Comparing price variation in different wholesale markets, for the New York City hub. The top graph shows a period when prices were similar across all markets; the bottom graph shows a period when there was significantly more volatility in the real-time market.

Surprisingly, negative prices can show up for brief periods, representing conditions where if energy were to be consumed at a specific location at a specific time the overall efficiency of the system would increase.

Market boundaries introduce economic transaction inefficiencies. As we shall see later, even geographically close locations in different markets tend to see uncorrelated prices. Part of the problem is that different markets have evolved using different rules, pricing models, etc.

Clearly, the market for electricity is complex. In addition to the factors mentioned here, many local idiosyncrasies exist. In this paper, we use a relatively simple market model that assumes the following:

1. Real-time prices are known and vary hourly.
2. The electric bill paid by the service operator is proportional to consumption and indexed to wholesale prices.
3. The request routing behavior induced by our method does not significantly alter prices and market behavior.

The validity of the second assumption depends upon the extent to which companies hedge their energy costs by contractually locking in fixed pricing (see section 7). A large body of economic literature deals with the structure and evolution of energy markets [19, 20, 21], market failures, and arbitrage opportunities for securities traders (e.g. [22, 23]).

3. EMPIRICAL MARKET ANALYSIS

We posit that imperfectly correlated variations in local electricity prices can be exploited by operators of large geographically distributed systems to save money. Rather than presenting a theoretical discussion, we take an empirical approach, grounding our analysis in historical market data aggregated from government sources [19, 16], trade publication archives [18], and public data archives maintained by the different RTOs. We use price data for 30 locations, covering January 2006 through March 2009.

3.1 Price Variation

Geographic price differentials are what really matter to us, but it is useful to first get a feel for the behaviour of individual prices.

Daily Variation. Figure 3 shows daily average prices for four locations⁶, from January 2006 through April 2009. Although prices are relatively stable at long time scales, they exhibit a significant amount of day-to-day volatility, short-term spikes, seasonal trends, and dependencies on fuel prices and consumer demand. Some locations in the figure are visibly correlated, but hourly prices are not correlated (§3.2).

Different Market Types. Spot and futures markets have different price dynamics. Figures 4 and 5 illustrate the difference for NYC. Compared to the day-ahead market, the hourly real-time (RT) market is more volatile, with more high-frequency variation, and a lower average price. The underlying five minute RT prices are even more volatile.

⁶The Northwest is an important region, but lacks an hourly wholesale market, forcing us to omit the region from the remainder of our analysis.
3.2 Geographic Correlation

Our approach would fail if hourly prices are well correlated at different locations. However, we find that locations in different regional markets are never highly correlated, even when nearby, and that locations in the same region are not always well correlated.

Figure 8 shows a scatter-plot of pairwise correlation and geographic distance. No pairs were negatively correlated. Note how correlation decreases with distance. Further, note the impact of RTO market boundaries: most pairs drawn from the same RTO lie above the 0.6 correlation line, while all pairs from different regions lie below it. We also see a surprising lack of diversity within some regions: LA and Palo Alto have a coefficient of 0.94.

Hourly prices are not correlated at short time-scales, but we should not expect prices to be independent. Natural gas prices, for example, will introduce some coupling (see figure 3) between distant locations.

3.3 Price Differentials

Figure 9 shows hourly price differentials for two pairs of locations over an eight day period (both pairs have mean differentials close to zero). The three locations are far from each other and in different RTOs. We see price spikes (some extend far off the scale, the largest is $1900) and extended periods of price asymmetry. Sometimes the asymmetry favours one, sometimes the other. This suggests that a pre-determined assignment of clients to servers is not optimal.

Differential Distributions. Consider two locations. In order for our dynamic approach to yield substantial savings over a static solution, the price differential between those locations must vary in time, and the distribution of this differential should ideally have a zero mean and a reasonably high variance. Such a distribution would imply that neither site is strictly better than the other, but also that a dynamic solution, always buying from whichever site is least expensive that hour, could yield meaningful savings. Additionally, the dynamic approach could win when presented with two locations having uncorrelated periods of price elevation.
Figure 10: Price differential histograms for five location pairs and 39 months of hourly prices.

Figure 11: PaloAlto-Virginia price differential distributions for each month. The monthly median prices and inter-quartile range are shown.

Figure 12: Price differential distributions (median and inter-quartile range) for each hour of the day.

Figure 13: For PaloAlto-Virginia, short-lived price differentials account for most of the time.

4. AKAMAI: TRAFFIC AND BANDWIDTH

In order to understand the interaction of real workloads with electricity prices, we acquired a data set detailing traffic on Akamai’s infrastructure. The data covers 24 days worth of traffic on a large subset of Akamai’s servers, with a peak of over 2 million hits/sec (figure 14). The 9-region traffic is the subset of servers for which we have electricity price data.

We use the Akamai traffic because it is a realistic workload. Akamai has over 2000 content provider customers in the US. Hence, the traffic represents a broad user base. There are 60 other pairs (a set of 16 hubs) with |µ| ≤ 5 ∧ σ ≥ 50; and 86 pairs (a set of 28 hubs) with |µ| ≤ 5 ∧ σ ≥ 25.
However, Akamai does not use aggressive server power management, their CDN is sensitive to latency and their workload contains a large fraction of computationally trivial hits (e.g., fetches of well cached objects). So our work is far less relevant to Akamai than to systems where more energy elasticity exists and workloads are computationally intensive. Furthermore, in mapping clients to servers, Akamai’s system balances a number of concerns—trying to optimize performance, handle partially replicated CDN objects, optimize network bandwidth costs, etc.

**Traffic Data.** Traffic data was collected at 5-minute intervals on servers housed in Akamai’s public clusters. Akamai has two types of clusters: public, and private. Private clusters are typically located inside of universities, large companies, small ISPs, and ISPs outside the US. These clusters are dedicated to serving a specific user base, e.g., the members of a university community, and no others. Public clusters are generally located in commercial co-location centers and can serve any users worldwide. For any user not served by a private cluster, Akamai has the freedom to choose which of its public clusters to direct the user. Clients that end up at public clusters tend to see longer network paths than clients that can be served at private clusters.

The 5-minute data contains, for each public cluster: the number of hits and bytes served to clients; a rough geography of where those clients originated; and the load in each of the clusters. In addition, we surveyed the hardware used in the different clusters and collected values for observed server power usage. We also looked at the top-level mapping system to see how name-servers were mapped to clusters.

In the data we collected, the geographic localization of clients is coarse: they are mapped to states in the US, or countries. If multiple clusters exist in a city, we aggregate them together and treat them as a single cluster. This affects our calculation of client-server distances in \[6\].

**Bandwidth Costs.** An important contributor to data center costs is bandwidth, and there may be large differences between costs on different networks, and sometimes on the same network over time. Bandwidth costs are significant for Akamai, and thus their system is aggressively optimized to reduce bandwidth costs. We note that changing Akamai’s current assignments of clients to clusters to reduce energy costs could increase its bandwidth costs (since they have been optimized already). Right now the portion of colocation cost attributable to energy is less than but still a significant fraction of the cost of bandwidth. The relative cost of energy versus bandwidth has been rising. This is primarily due to decreases in bandwidth costs.

We cannot ignore bandwidth costs in our analysis. The complication is that the bandwidth pricing specifics are considered to be proprietary information. Therefore, our treatment of bandwidth costs in this paper will be relatively abstract.

Akamai does not view bandwidth prices as being geographically differentiated. In some instances, a company as large as Akamai can negotiate contracts with carriers on a nationwide basis. Smaller regional providers may provide transit for free. Prices are usually set per network port, using the basic 95/5 billing model: traffic is divided into five minute intervals and the 95th percentile is used for billing.

Our approach in this paper is to estimate 95th percentiles from the traffic data, and then to constrain our energy-price rerouting so that it does not increase the 95th percentile bandwidth for any location.

**Client-Server Distances.** Lacking any network level data on clients, we use geographic distance as a coarse proxy for network performance in our simulations. We see some evidence of geo-locality in the Akamai traffic data, but there are many cases where clients are not mapped to the nearest cluster geographically. One reason is that geographical distance does not always correspond to optimal network performance. Another possibility is that the system is trying to keep those clients on the same network, even if Akamai’s servers on that network are geographically far away. Yet another possibility is that clients are being moved to distant clusters because of 95/5 bandwidth constraints.

### 5. Modeling Energy Consumption

In order to estimate by how much we can reduce energy costs, we must first model the system’s energy consumption for each cluster. We use data from the Akamai CDN as a representative real-world workload. This data is used to derive a distribution of client activity, cluster sizes, and cluster locations. We then use an energy model to map prices and cluster-trafﬁc allocations to electricity expenses. The model is admittedly simplistic. Our goal is not to provide accurate ﬁgures, but rather to estimate bounds on savings.

#### 5.1 Cluster Energy Consumption

We model the energy consumption of a cluster as being proportional, roughly linear, to its utilization. Multiple studies have shown that CPU utilization is a good estimator for power usage \([5, 8]\). Our model is adapted from Google’s empirical study of a data center \([5]\) in which their model was found to accurately (less than 1% error) predict the dynamic power drawn by a group of machines (20-60 racks).
We augment this model to fill in some missing pieces and parameterize it using other published studies and measurements of servers at Akamai.

Let $P_{\text{cluster}}$ be the power usage of a cluster, and let $u_t$ be its average CPU utilization (between 0 and 1) at time $t$:

$$P_{\text{cluster}}(u_t) = F(n) + V(u_t, n) + \epsilon$$

Where $n$ is the number of servers in the cluster, $F$ is the fixed power, $V$ is the variable power, and $\epsilon$ is an empirically derived correction constant (see [5]).

$$F(n) = n \cdot (P_{\text{idle}} + (PUE - 1) \cdot P_{\text{peak}})$$

$$V(u_t, n) = n \cdot (P_{\text{peak}} - P_{\text{idle}}) \cdot (2u_t - u_t^r)$$

Where $P_{\text{idle}}$ is the average idle power draw of a single server, $P_{\text{peak}}$ is the average peak power, and the exponent $r$ is an empirically derived constant equal to 1.4 (see [5]). The equation for $V$ is taken directly from the original paper. A linear model ($r = 1$) was also found to be reasonably accurate [5]. We added the PUE component, since the Google study did not account for cooling etc.

With power-management, the idle power consumption of a server can be as low as 50-65% of the peak power consumption, which can range from 100-250W [5, 7, 8]. Without power-management an off-the-shelf server purchased in the last several years averages around 250W and draws $\sim$95% of its peak power when idle (based on measured values).

Ultimately, we want to use this model in simulation to estimate the maximum percentage reduction in the energy costs of some server deployment pattern. Consequently, the absolute values chosen for $P_{\text{peak}}$ and $P_{\text{idle}}$ are unimportant: their ratio is what matters. In fact, it turns out that the value $P_{\text{cluster}}(0) / P_{\text{cluster}}(1)$ is critical in determining the savings that can be achieved using price-differential aware routing.

Ideally, $P_{\text{cluster}}(0)$ would be zero: an idle cluster would consume no energy. At present, achieving this without impacting performance is still an open challenge. However, there is an increasing interest in energy-proportional computing [6] and dynamic server provisioning techniques are being explored by both academics and industry [7, 8, 9, 10, 11]. We are confident that $P_{\text{cluster}}(0)$ will continue to fall.

### 5.2 Increase in Routing Energy

In our scheme, clients may be routed to distant servers in search of cheap energy. From an energy perspective, this network path expansion represents additional work that must be performed by something. If this increase in energy was significant, network providers might attempt to pass the additional cost on to the server operators. Given what we know about bandwidth pricing (§4), a small increase in routing energy should not impact bandwidth prices. Alternatively, server operators may bear all the increased energy costs (suppose they run the intermediate routers).

A simple analysis suggests that the increased path lengths will not significantly alter energy consumption. Routers are not designed to be energy proportional and the energy used by a packet to transit a router is many orders of magnitude below the energy expended at the endpoints (e.g., Google’s 1 kJ/query [24]). We estimate that the average energy needed for a packet to pass through a core router is on the order of 2 mJ [25]. Further we estimate that the incremental energy dissipated by each packet passing through a core router would be as low as a 50 μJ per medium-sized packet [25].

We must also consider what happens if the new routes overload existing routers. If we use enough additional bandwidth through a router it may have to be upgraded to higher capacity hardware, increasing the energy significantly. However, we could prevent this by incorporating constraints, like the 95/5 bandwidth constraints we use.

### 6. SIMULATION: PROJECTING SAVINGS

In order to test the central thesis of this paper, we conducted a number of simulations, quantifying and analyzing the impact of different routing policies on energy costs and client-server distance.

Our results show that electricity costs can plausibly be reduced by up to 40% and that the degree of savings primarily depends on the energy elasticity of the system, in addition to bandwidth and performance constraints. We simulate Akamai’s 95/5 bandwidth constraints and show that overall system costs can be reduced. We also sketch the relationship between client-server distance and savings. Finally we investigate how delaying the system’s reaction to price differentials affects savings.

#### 6.1 Simulation Strategy

We constructed a simple discrete time simulator that stepped through the Akamai usage statistics, letting a routing module (with a global view of the network) allocate traffic to clusters at each time step. Using these allocations, we modeled each cluster’s energy consumption, and used observed hourly market prices to calculate energy expenditures. Before presenting the results, we provide some details about our simulation setup.

**Electricity Prices.** We used hourly real-time market prices for twenty-nine different locations (hubs). However, we only have traffic data for Akamai public clusters in nine of these locations. Therefore, most of the simulations focused on these nine locations. Our data set contained 39 months of price data, spanning January 2006 through March 2009. Unless noted otherwise, we assumed the system reacted to the previous hour’s prices.

**Traffic and Server Data.** The Akamai workload data set contains 5-minute samples for the hits-per-second observed at public clusters in twenty five cities, for a period of 24 days and some hours. Each sample also provides a map, specifying where hits originated, grouping clients by state, and which city they were routed to.

We had to discard seven of these cities because of a lack of electricity market data for them. The remaining eighteen cities were grouped by electricity market hub, as nine ‘clusters’. In our 24-day simulation, we used the traffic incident on these nine clusters.

In order to simulate longer periods we derived a synthetic workload from the 24-day Akamai workload (US traffic only). We calculated an average hit rate For every hub and client state pair. We produced a different average for each hour of the day and each day of the week.

Additionally, the Akamai data allowed us to derive capac-

---

11Reported: power consumption of idle router is 97% the peak power. In the future, power-aware hardware may reduce this disparity between the marginal and average energy.

---
Figure 15: The system’s energy elasticity is key in determining the degree of savings price-conscious routing can achieve. Further, obeying existing 95/5 bandwidth constraints reduces, but does not eliminate savings. The graph shows 24-day savings for a number of different PUE and \( P_{idle} \) values with a 1500km distance threshold. The savings for each energy model are given as a percentage of the total electricity cost of running Akamai’s actual routing scheme under that energy model.

Routing Schemes. In our simulations we look at two routing schemes: Akamai’s original allocation; and a distance-constrained electricity price optimizer.

Given a client, the price-conscious optimizer maps it to a cluster with the lowest price, only considering clusters within some maximum radial geographic distance. For clients that do not have any clusters within that maximum distance, the routing scheme finds the closest cluster and considers any other nearby clusters (< 50km). If the selected cluster is nearer its capacity (or the 95/5 boundary), the optimizer iteratively finds another good cluster.

The price optimizer has two parameters that modulate its behaviour: a distance threshold and a price threshold. Any price differentials smaller than the price threshold are ignored (we use $5/MWh). Setting the distance threshold to zero, gives an optimal distance scheme (select the cluster geographically closest to client); setting it to a value larger than the East-West coast distance gives an optimal price scheme (always select the cluster with the lowest price).

We are not proposing this as a candidate for implementation, but it allows us to benchmark how well a price-conscious scheme could do and to investigate trade-offs between distance constraints and achievable savings.

Energy Model. We use the cluster energy model from section 5.1. We simulated the running cost of the system using a number of different values for the peak server power (\( P_{peak} \)), idle server power (\( P_{idle} \)) and the PUE. This section discusses normalized costs and \( P_{idle} \) is always expressed as a percentage of \( P_{peak} \). Some energy parameters that we used: optimistic future (0% idle, 1.1 PUE); cutting-edge/google (60% idle, 1.3 PUE); state-of-the-art (65% idle, 1.7 PUE); disabled power management (95% idle, 2.0 PUE).

Client-Server Distance. Given a client’s origin state and the server’s location (hub), our distance metric calculates a population-density weighted geographic distance. We used census data to derive basic population density functions for each US state. When the traffic contains clients from outside the US, we ignore them in the distance calculations.

We use this function as a coarse measure for network distance. The granularity of the Akamai data set does not provide enough information for us to estimate network latency between clients and servers, or even to accurately calculate geographic distances between clients and servers.

6.2 At the Turn of the Year: 24 Days of Traffic

We begin by asking the question: what would have happened if an Akamai-like system had used price conscious routing at the end of 2008? How would this have compared in cost and client-server distance to the current routing methods employed by Akamai?

Energy Elasticity. We find that the answer hinges on the energy elasticity characteristics of the system. Figure 15 illustrates this. When consumption is completely proportional to load, using price-conscious routing could eliminate 40% of the electricity expenditure of Akamai’s traffic allocation, without appreciably increasing client-server distances. As idle server power and PUE rise, we see a dramatic drop in possible savings: at Google’s published elasticity level (65%, 1.3 PUE), the maximum savings have dropped to 5%. Inelasticity constrains our ability to route power demand away from high prices.

Bandwidth Costs. A reduced electric bill may be overshadowed by increased bandwidth costs. Figure 15 therefore also shows the savings when we prevent clusters from having higher 95th percentile hit rates than were observed in the Akamai data. We see that constraining bandwidth in this way may cause energy savings to drop down to about a third of their earlier values. However, the good news is that these savings are reductions in the total operating cost.

By jointly optimizing bandwidth and electricity, it should be possible to acquire part of the economic value represented by the difference between savings with and without bandwidth constraints.

Distance and Savings. The savings in figure 15 do not represent a free lunch: the mean client-server distance may need to increase to leverage market diversity.

The price conscious routing scheme we use has a distance threshold parameter, allowing us to explore how higher client-server distances lead to lower electric bills. Figure 16 shows how increasing the distance threshold can be used to reduce electricity costs. Figure 17 shows how client-server distances change in response to changes in the threshold.

At a distance threshold of 1100km, the 99th percentile estimated client-server distances is at most 800km. This should provide an acceptable level of performance (the distance between Boston and Alexandria in Virginia is about 650km and network RTTs are around 20ms).

At this threshold, using the future energy model, the savings is significant, between 10% (obey 95/5 constraints) and 20%. There is an elbow at a threshold of 1500km, causing both savings and distances to jump (the distance between Boston and Chicago is about 1400km). After this, increasing the threshold provides diminishing returns.
Figure 16: 24-day electricity costs fall as the distance threshold is increased. The costs shown here are for a (0% idle, 1.1 PUE) model, normalized to the cost of the Akamai allocation.

Figure 17: Increasing the distance threshold allows the routing of clients to cheaper clusters further away (figure 16 shows corresponding falling cost).

Figure 18: 39-month electricity costs fall as the distance threshold is increased. The costs shown here are for a (0% idle, 1.1 PUE) model, normalized to the cost of the synthetic Akamai-like allocation.

Figure 19: Change in per-cluster cost for 39-month simulations with different distance thresholds. This uses the future (0%, 1.1) model, and obeys 95/5 constraints.

### Dynamic Beats Static

Dynamic Beats Static. In particular, we see that when 95/5 constraints are ignored, the dynamic cost minimization solution can be substantially better than a static one. In figure 18, we see that the dynamic solution could reduce the electricity cost down to almost 55%, while moving all the servers to the region with the lowest average price would only reduce cost down to 65%.

### 6.3 Synthetic Workload: 39 Months of Prices

The previous section uses a very small subset of the price data we have. Using a synthetic workload, derived from the original 24-day one, we ran simulations covering January 2006 through March 2009. Our results show that savings increase above those for the 24-day period.

Figure 18 shows how electricity cost varied with the distance threshold (analogous to figure 16). The results are similar to what we saw for the 24-day case, but maximum savings are higher. Notably: thresholds above 2000km in figure 18 do not exhibit sharply diminishing returns like those seen in 16. In order to normalize prices, we used statistics of how Akamai routed clients to model an Akamai-like router, and calculated its 39-month cost.

Figure 19 breaks down the savings by cluster, showing the change in cost for each cluster. The largest savings is shown at NYC. This is not surprising since the highest peak prices tend to be in NYC. These savings are not achieved by always routing requests away from NYC: the likelihood of requests being routed to NYC depends on the time of day.

We simulated other server distributions (evenly distributed across all 29 hubs, heterogeneous distributions, etc) and saw similar decreasing cost/distance curves.

Dynamic Beats Static. In particular, we see that when 95/5 constraints are ignored, the dynamic cost minimization solution can be substantially better than a static one. In figure 18, we see that the dynamic solution could reduce the electricity cost down to almost 55%, while moving all the servers to the region with the lowest average price would only reduce cost down to 65%.

### 6.4 Reaction Delays

Not reacting immediately to price changes noticeably reduces overall savings. In our simulations we were conservative and assumed that there was a one hour delay between the market setting new prices and the system propagating new routes.

Figure 20 shows how increasing the reaction delay impacts prices. First, note the initial jump, between an immediate reaction and a next-hour reaction. This implies achievable savings will exceed what we have calculated for systems that can update their routes in less than an hour. Further, note the local minima at the 24 hour mark. This is probably because market prices can be correlated for a given hour from one day to the next.

The increase in cost is substantial. With the (65% idle, 1.3 PUE) energy model, the maximum savings is around 5% (see figure 15). So a subsequent increase in cost of 1% would eliminate a large chunk of the savings.
Commonwealth Edison offers a pricing structure that appeals to electricity providers since risk is transferred to consumers. For example, in the mid-west RTO market to take advantage of our techniques. This billing structure appeals to electricity providers since risk is transferred to consumers. For example, in the mid-west RTO Commonwealth Edison offers a Real-Time Pricing program [26]. Customers enrolled in it are billed based on hourly consumption and corresponding wholesale PJM-MISO local market prices.

Companies, such as Akamai, renting space in co-location facilities will almost certainly have to negotiate a new billing structure to get any advantage from our approach. Most co-location centers charge by the rack, each rack having a maximum power rating. In other words, a company like Akamai pays for provisioned power, and not for actual power used. We speculate that as energy costs rise relative to other costs, it will be in the interest of co-location owners to charge based on consumption and possibly location. There is evidence that bandwidth costs are falling, but energy costs are not. Even if new kinds of contracts do not arise, server operators may be able to sell their load-flexibility through a side-channel like demand response, as discussed below, bypassing inflexible contracts.

Selling Flexibility. Distributed systems with energy elastic clusters can be more flexible than traditional consumers: operators can quickly and precipitously reduce power usage at a location (by suspending servers, and routing requests elsewhere). Market mechanisms already exist that would allow operators to value and sell this flexibility. Some RTOs allow energy users to bid negawatts (negative demand, or load reductions) into the day-ahead market auction. This is believed to moderate prices.

Alternatively, customers could enroll in triggered demand response programs, agreeing to reduce their power usage in response to a request by the grid operators. Load reduction requests are sent out when electricity demand is high enough to put grid reliability at risk, or rising demand requires the imminent activation of expensive/unreliable generation assets. The advance notice given by the RTO can range from days to minutes. Participating customers are compensated based on their flexibility and load. Demand-response variants exist in every market we cover in this paper.

Even consumers using as little as 10kW (a few racks) can participate in such programs. Consumers can also be aggregated into large blocs that reduce load in concert. This is the approach taken by EnerNOC, a company that collects many consumers, packages them, and sells their aggregate ability to make on-demand reductions. A package of hotels would, for example, reduce laundry volume in sync to ease power demand on the grid.

The good thing about selling flexibility as a product, is that this is valued even where wholesale markets do not exist. It even works if price-differentials don’t exist (e.g. fixed price contracts or in highly regulated markets).

However, we have ignored the demand side. How do operators construct bids for the day-ahead auctions if they don’t know next-day client demand for each region? What happens when operators are told to reduce power consumption at a location, when there is a concentration of active clients nearby? In systems like Akamai, demand is generally predictable, but there will be heavy traffic days that are impossible to predict.

There is anecdotal evidence that data centers have participated in demand response programs [3]. However, the applicability of demand response to single data centers is not widely accepted. Participating data centers may face additional downtime or periods of reduced capacity. Conversely, when we look at large distributed systems, participation in such programs is attractive. Especially when the barriers to entry are so low—only a few racks per location are needed to construct a multi-market demand response system.

8. FUTURE WORK

Some clear avenues for future work exist.

Implementing Joint Optimization. Existing systems already have frameworks in place that engineer traffic to optimize for bandwidth costs, performance, and reliability. Dynamic energy costs represent another input that should be integrated into such frameworks.

RTO Interaction. Service operators can interact with RTOs in many ways. This paper has proposed a relatively passive approach in which operators monitor spot prices and react to favourable conditions. As we discussed in section 7, there are other market mechanisms in place that service operators may be able to exploit. The optimal market participation strategy is unclear.

Weather Differentials. Data centers expend a lot of energy running air cooling systems, up to 25% of total energy. In modern systems, when ambient temperatures are low enough, external air can be used to radically reduce the power draw of the chillers. At the same time, weather temperature differentials are common. This suggests that significant energy savings can be achieved by dynamically
routing requests to sites where the heat generated by serving the request is most inexpensively removed. Unlike price differentials, which reduce cost but not energy, routing requests to cooler regions may be able to reduce both.

Environmental Cost. Rather than attempting to minimize the dollar cost of the energy consumed, a socially responsible service operator may instead choose to use an environmental impact cost function. The environmental impact of a service is time-varying. An obvious cost function is the carbon footprint of the energy used. In grids that aggregate electricity from diverse providers, the footprint varies depending upon what generating assets are active, whether power plants are operating near optimal capacity and what mixture of fuels they are currently using. The variation occurs at multiple time scales, e.g., seasonal (is there water to power hydro systems), weekly (what are the relative prices of various fossil fuels), and hourly (is the wind blowing or the tide going out). Additionally, carbon is not the only pollutant. For instance, power plants are the primary stationary sources of nitrogen oxide in the US. Due to variations in weather and atmospheric chemistry, the timing and location of NOx reductions determine their effectiveness in reducing ground-level ozone [27].

9. CONCLUSION

The bounds derived in this paper should not be taken too literally. Our cost and traffic models are based on actual data, but they do incorporate a number of simplifying assumptions. The most relevant assumptions are probably (1) that operators can do better by buying electricity on the open market than through negotiated long-term contracts, and (2) that the variable energy costs associated with servicing a request are a significant fraction of the total costs.

Despite these caveats, it seems clear that the nature of geographical and temporal differences in the price of electricity offers operators of large distributed systems an opportunity to reduce the cost of servicing requests. It should be possible to augment existing optimization frameworks to deal with electricity prices.

Acknowledgements

We thank our shepherd Jon Crowcroft and the anonymous reviewers for their insightful comments. We also thank John Parsons, Ignacio Perez-Arriaga, Harihara Shankar Rahul, and Noam Freedman for their help. This work was supported in part by Nokia, and by the National Science Foundation under grant CNF-0435382.

10. REFERENCES