Cutting the Electric Bill for Internet-Scale Systems

Asfandyar Qureshi
MIT CSAIL
asfandyar@mit.edu

Rick Weber
Akamai Technologies
riweber@akamai.com

Hari Balakrishnan
MIT CSAIL
hari@mit.edu

John Guttag
MIT CSAIL
guttag@mit.edu

Bruce Maggs
Carnegie Mellon University
bmm@cs.cmu.edu

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.

SIGCOMM'09, August 17–21, 2009, Barcelona, Spain.

Company	Servers	Electricity	Cost
eBay	16K	$\sim 0.6\times10^5$ MWh	~$3.7M
Akamai	40K	$\sim 1.7\times10^5$ MWh	~$10M
Rackspace	50K	$\sim 2\times10^5$ MWh	~$12M
Microsoft	>200K	$>6\times10^5$ MWh	>$36M
Google	>500K	$>6.3\times10^5$ MWh	>$38M
USA (2006)	10.9M	610×10^5 MWh	$4.5B
MIT campus		2.7×10^5MWh	$62M

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.

Organizations such as Google, Microsoft, Amazon, Yahoo!, and many other operators of large networked 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[1] consumes an estimated $10M worth of electricity annually[2].

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, multicore 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

[1]This paper covers work done outside Akamai and does not represent the official views of the company.
[2]Though Akamai seldom pays directly for electricity, it pays for it indirectly as part of co-location expenses.

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 2011, data center energy use could double. At best, 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[3]. The

[3]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.

RTO	Region	Some Regional Hubs
ISONE	New England	Boston (MA-BOS), Maine (ME), Connecticut (CT)
NYISO	New York	NYC, Albany (CAPITL), Buffalo (WEST), PJM import (PJM)
PJM	Eastern	Chicago (CHI), Virgina (DOM), New Jersey (NJ)
MISO	Midwest	Peoria (IL), Minnesota (MN), Indiana (CINERGY)
CAISO	California	Palo Alto (NP15), LA (SP15)
ERCOT	Texas	Dallas (N), Austin (S)

Figure 2: The different regions studied in this paper. The listed hubs provide a sense of RTO coverage and a reference to map electricity market location identifiers (hub NP15) to real locations (Palo Alto).

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)[4] 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 6.3×10^5 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 1×10^5 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].

[4]A measure of data center energy efficiency.

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* electricty 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[5].

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

[5]Hour-ahead markets, not discussed here, are analogous.

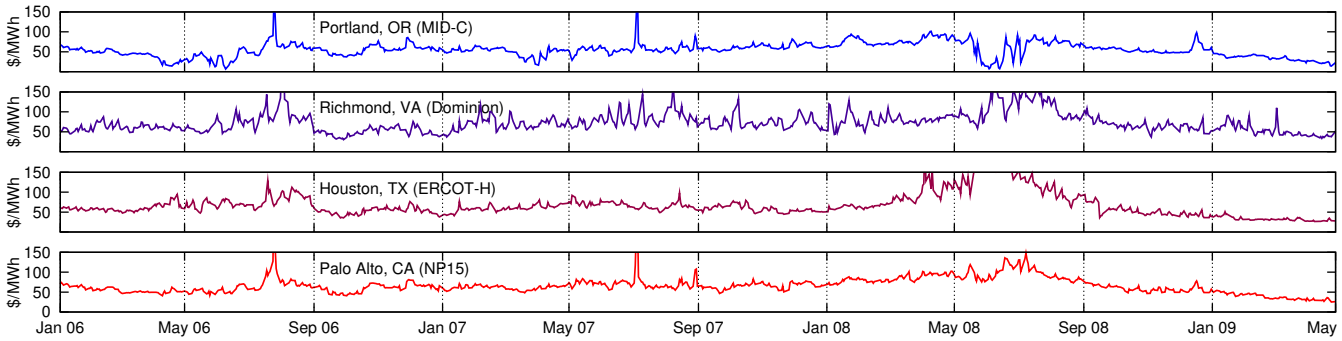


Figure 3: Daily averages of day-ahead peak prices at different hubs [18]. The elevation in 2008 correlates with record high natural gas prices, and does not affect the hydroelectric dominated Northwest. The Northwest consistently experiences dips near April (this seems to be correlated with seasonal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

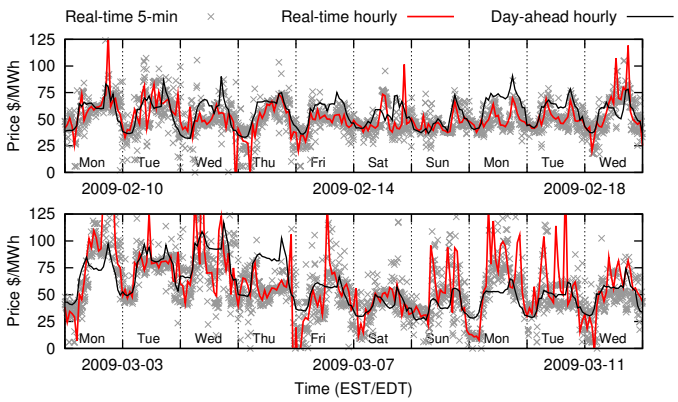


Figure 4: Comparing price variation in different wholesale markets, for the New York City hub. The top graph shows a period when prices were similar across all markets; the bottom graph shows a period when there was significantly more volatility in the real-time market.

expensive generation units will then need to be activated, driving up prices. Some markets include an explicit congestion cost component in their prices.

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

Window	5 min	1 hr	3 hr	12 hr	24 hr
Real-time σ	28.5	24.8	21.9	18.1	15.6
Day-ahead σ	N/A	20.0	19.4	17.1	16.0

Figure 5: The real-time market is more variable at short time-scales than the day-ahead market. Standard deviations for Q1 2009 prices at the NYC hub are shown, averaged using different window sizes.

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.

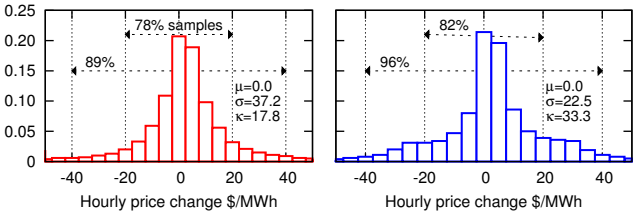
Daily Variation. Figure 3 shows daily average prices for four locations[6], from January 2006 through April 2009. Although prices are relatively stable at long time scales, they exhibit a significant amount of day-to-day volatility, short-term spikes, seasonal trends, and dependencies on fuel prices 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.

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

Location	RTO	Mean*	StDev*	Kurt.*
Chicago, IL	PJM	40.6	26.9	4.6
Indianapolis, IN	MISO	44.0	28.3	5.8
Palo Alto, CA	CAISO	54.0	34.2	11.9
Richmond, VA	PJM	57.8	39.2	6.6
Boston, MA	ISONE	66.5	25.8	5.7
New York, NY	NYISO	77.9	40.26	7.9

Figure 6: Real-time market statistics, covering hourly prices from January 2006 through March 2009 (*statistics are from the 1% trimmed data).



(a) Palo Alto (NP15) **(b) Chicago (PJM)**

Figure 7: Histograms of hour-to-hour change in real-time hourly prices for two locations, over the 39-month period. Both distributions are zero-mean, Gaussian-like, with very long tails.

For the remainder of this paper, we focus exclusively on the RT market. Our goal is to exploit geographically uncorrelated volatility, something that is more common in the RT market. We restrict ourselves to hourly prices, but speculate that the additional volatility in five minute prices provides further opportunities.

Figure 6 provides additional statistics for hourly RT prices.

Hour-to-Hour Volatility. As seen in figure 4, the hour-to-hour variation in NYC's RT prices can be dramatic. Figure 7 shows the distribution of the hourly change for Palo Alto and Chicago. At each location, the price per MWh changed hourly by $20 or more roughly 20% of the time. A $20 step represents 50% of the mean price for Chicago. Furthermore, the minimum and maximum price during a single day can easily differ by a factor of 2.

The existence of rapid price fluctuations reflects the fact that short term demand for electricity is far more elastic than supply. Electricity cannot always be efficiently moved from low demand areas to high demand areas, and producers cannot always ramp up or down easily.

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[7]. 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[8]. We also see

[7]We have verified our results using subsets of the data (e.g. last 12 months), *mutual information* ($I_{x,y}$), shifted signals, etc.

[8]$I_{x,y}$ much more clearly divides the data between same-RTO and different-RTO pairs, suggesting that the small overlap in figure 8 is due to the existence of non-linear relationships within NYISO and ERCOT, not detected by the correlation coefficient.

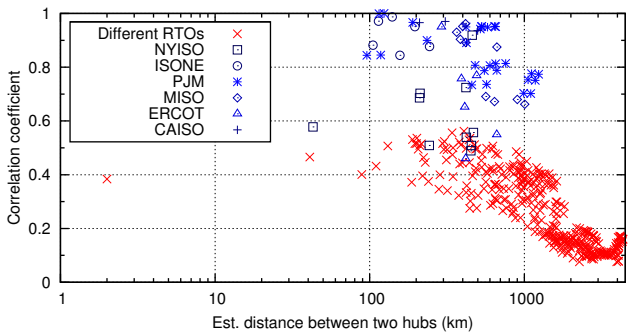


Figure 8: The relationship between price correlation, distance, and parent RTO. Each point represents a pair of hubs (29 hubs, 406 pairs), and the correlation coefficient of their 2006-2009 hourly prices ($> 28k$ samples each). Red points represent paired hubs from different RTOs; blue points are labelled with the RTO of both.

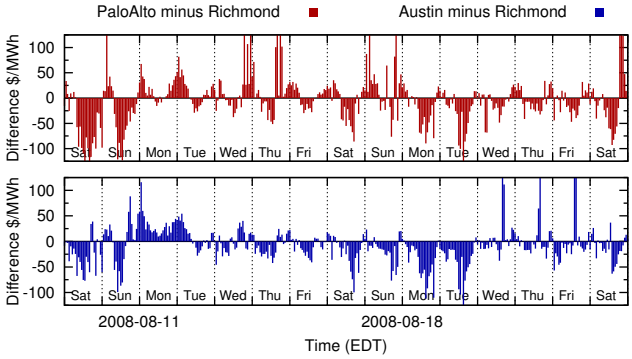


Figure 9: Variation of price differentials with time.

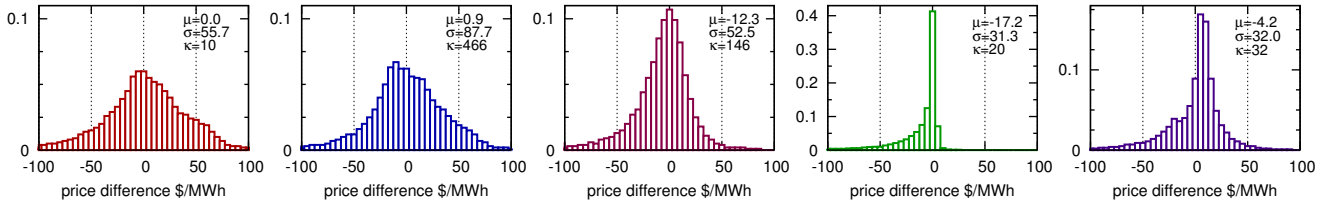
a surprising lack of diversity within some regions: LA and Palo Alto have a coefficient of 0.94.

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.



(a) PaloAlto - Virginia (b) Austin - Virginia (c) Boston - NYC (d) Chicago - Virginia (e) Chicago - Peoria

Figure 10: Price differential histograms for five location pairs and 39 months of hourly prices.

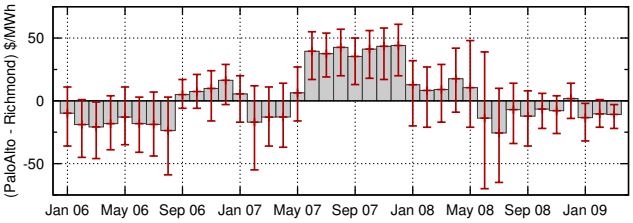


Figure 11: PaloAlto-Virginia price differential distributions for each month. The monthly median prices and inter-quartile range are shown.

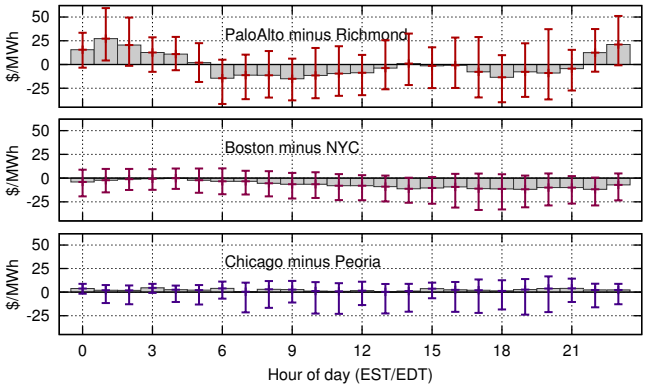


Figure 12: Price differential distributions (median and inter-quartile range) for each hour of the day.

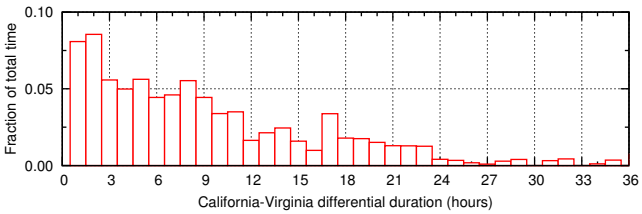


Figure 13: For PaloAlto-Virginia, short-lived price differentials account for most of the time.

Figure 10 shows the pairwise differential distributions for some locations, for the 2006-2009 data. The California-Virginia (figure 10a) and Texas-Virginia (figure 10b) distributions are zero-mean with a high variance. There are many other such pairs[9].

Boston-NYC (figure 10c) is skewed, since Boston tends to be cheaper than NYC, but NYC is less expensive 36% of the time (the savings are greater than $10/MWh 18% of the time). Thus, even with such a skewed distribution, there exists an opportunity to dynamically exploit differentials for meaningful savings.

Unsurprisingly, a number of pairs exist where one location is strictly better than the other, and dynamic adaptation is unnecessary. Chicago-Virginia (figure 10d) is an example: Virginia is less expensive 8% of the time, but the savings almost never exceed $10/MWh.

The dispersion introduced by a market boundary can be seen in the dynamically exploitable Chicago-Peoria distribution (figure 10e).

Evolution in Time. The price differential distributions do not remain static in time. Figure 11 shows how the PaloAlto-Virginia distribution changed from month to month. A sustained price asymmetry may exist for many months, before reversing itself. The spread of prices in one month may double the next month.

Time-of-Day Price differentials depend on the time-of-day. For instance, because California and Virginia are in different time zones, peak demand does not overlap. This is likely an important factor shaping the price differential.

Figure 12 shows how the hour of day affects the differentials for three location pairs. For PaloAlto-Virginia, we see a strong dependency on the hour. Before 5am (eastern), Virginia has a significant edge; by 6am the situation has reversed; from 1-4pm neither is better. For Boston-NYC we see a different kind of dependency: from 1am-7am neither site is better, at all other times Boston has the edge. The effect of hour-of-day on Chicago-Peoria is less clear.

Differential Duration. We define the *duration* of a sustained price differential as the number of hours one location is favoured over another by more than $5/MWh. As soon as the differential falls below this threshold, or reverses to favour the other location, we mark the end of the differential.

Figure 13 shows how much time was spent in short-duration price-differentials, for PaloAlto-Virginia. Short differentials (<3 hrs) are more frequent than other types. Medium length differentials (<9 hrs) are common. Differentials that last longer than a day are rare, for a balanced pair like this.

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.

[9]There are 60 other pairs (a set of 16 hubs) with $|\mu| \leq 5 \wedge \sigma \geq 50$; and 86 pairs (a set of 28 hubs) with $|\mu| \leq 5 \wedge \sigma \geq 25$.

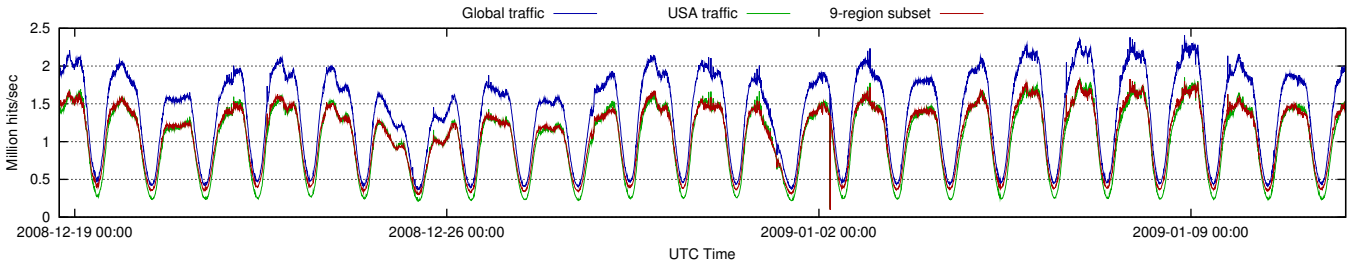


Figure 14: Traffic in the Akamai data set. We see a peak hit rate of over 2 million hits per second. Of this, about 1.25 million hits come from the US. The traffic in this data set comes from roughly half of the servers Akamai runs. In comparison, in total, Akamai sees around 275 billion hits/day.

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 world-wide. 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 co-location 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 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 parametrize it using other published studies and measurements of servers at Akamai.

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

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

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

$$\begin{aligned} F(n) &= n \cdot \left(P_{idle} + (PUE - 1) \cdot P_{peak}\right) \\ V(u_t, n) &= n \cdot (P_{peak} - P_{idle}) \cdot (2u_t - u_t^r) \end{aligned}$$

Where P_{idle} is the average idle power draw of a single server, P_{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 ∼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_{peak} and P_{idle} are unimportant: their ratio is what matters. In fact, it turns out that the value $\frac{P_{cluster}(0)}{P_{cluster}(1)}$ is critical in determining the savings that can be achieved using price-differential aware routing.

Ideally, $P_{cluster}(0)$ 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 $P_{cluster}(0)$ 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 were 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][10]. 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][11].

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 analysing 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-

[10] Reported for a Cisco GSR 12008 router: 540k mid-sized packets/sec and 770 Watts measured.

[11] Reported: 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.

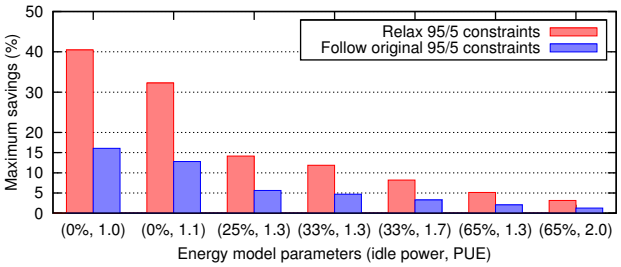


Figure 15: The system's energy elasticity is key in determining the degree of savings price-conscious routing can achieve. Further, obeying existing 95/5 bandwidth constraints reduces, but does not eliminate savings. The graph shows 24-day savings for a number of different PUE and P_{idle} values with 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.

ity constraints and the 95th percentile hits and bandwidth for each cluster. Capacity estimates were derived using observed hit rates and corresponding region load level data provided by Akamai. Our simulations use hits rather than the bandwidth numbers from the data.

Most of our simulations used Akamai's geographic server distribution. Although the details of the distribution may introduce artifacts into our results, this is a real-world distribution. As such, we feel relying on it rather than relying on synthetic distributions makes our results more compelling.

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 nearing 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% idle, 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.