# Green Enterprise Computing Data: Assumptions and Realities

Maria Kazandjieva, Brandon Heller, Omprakash Gnawali *, Philip Levis, and Christos Kozyrakis
Stanford University, *University of Houston
mariakaz, brandonh, pal, christos@cs.stanford.edu, * gnawali@cs.uh.edu

## Abstract

Until now, green computing research has largely relied on few, short-term power measurements to characterize the energy use of enterprise computing. This paper brings new and comprehensive power datasets through Powernet, a hybrid sensor network that monitors the power and utilization of the IT systems in a large academic building. Over more than two years, we have collected power data from 250+ individual computing devices and have monitored a subset of CPU and network loads. This dense, long-term monitoring allows us to extrapolate the data to a detailed breakdown of electricity use across the building's computing systems.

Our datasets provide an opportunity to examine assumptions commonly made in green computing. We show that power variability both between similar devices and over time for a single device can lead to cost or savings estimates that are off by 15-20%. Extending the coverage of measured devices and the duration (to at least one month) significantly reduces errors. Lastly, our experiences with collecting data and the subsequent analysis lead to a better understanding of how one should go about power characterization studies. We provide several methodology guidelines for future green computing research.

## I. Introduction

Common sense tells us that there are opportunities to reduce the energy waste of computing systems. For example, many people leave their computers on overnight, even when they are not needed, or have power-hungry PCs for undemanding tasks such as document-processing and web browsing. Observations like these have motivated recent research into green computing work [8], [13], [19].

Unfortunately, the data to support and evaluate new green computing solutions remains vastly anecdotal. Until now, power characterization studies have either collected data at the macro scale of a whole building [9], lumping all plug loads into one number, or at the micro scale from a handful of computers and LCD monitors [16]. Data at the macro scale is informative but difficult to act upon – it does not provide visibility into the computing components that can be made more energy efficient. Power data at the micro scale is great at providing a detailed characterization of single device but fail to show how the individual datapoint relates to the full building energy use.

The green computing research community can benefit from the availability of more extensive power measurements. For example, a single PC power measurement from 2004 [12] has been used in papers as recent as 2010, citing it as a representative value. The aforementioned paper gives the power draw of a 2002 Dell 2350 1.8 Ghz computer as 60 to 85 watts. A 2009 paper [7] measured two desktops (102 and 72 watts, respectively) and said their measurements were consistent with prior data, citing [12]. Later the same year, a characterization study [9] used 100 watts per desktop plus LCD for some of its calculations, citing [7]. In 2010, LiteGreen [13] also referenced [7], stating that the typical PC draws 80–100 watts when active. The paper goes on to measure one PC (95W active) and uses it to calculate potential energy savings of their proposed solution.

If we were to continue on citation trails like the one above, we risk using limited and possibly outdated data for new systems' evaluation. Fast-paced improvements in personal computing mean that some newer, more powerful PCs are also more power-hungry than the 60- to 100-watt range. In addition, enterprise environments are often heterogeneous and it is beneficial to have power measurements from a larger selection of devices.

This paper helps fill the power data gap by characterizing energy data at the individual– and the building–scale levels. Every month, Stanford's computer science department's pays a \$40,000 (330,000 kWh) electricity bill but there is no visibility into exactly where this energy is going and how much of it is spent on wasteful computing systems. This problem has much greater implications than a single department's budget. When one considers the problem at scale, computing infrastructures add up to billions of dollars. According to the latest Department of Energy "Annual Energy Review" [1], in 2003, computing in education and office buildings consumed 66 billion kilowatt-hours of electricity [2]. This is 2% of all US electricity consumption and, as with data centers, is increasing.

This paper presents Powernet, a multi-year study on the computing infrastructure in our department. Over two years, we have measured plug loads from over 250 devices as

[1]Released in October of 2011

well as utilization rates of a subset of the computers and networking equipment. The data from the deployment is available at **http://sing.stanford.edu/maria/powernet**. The measurement points were selected carefully instead of being a uniform, random sample. Since power does not necessarily follow a simple distribution over device type, outliers may be extremely important yet difficult to capture with a random sample.

Measuring over a long term allows us to quantify measurement errors over shorter durations. Measuring a large and diverse number of devices allows us to characterize the variation of power draw and utilization across and within device types. We augment our datasets with metadata including network device registrations and explicit equipment inventories. The combination of power data, utilization statistics, and metadata allows us to answer several open questions about green computing:

- What is the contribution of computing systems to an enterprise's overall electricity consumption and waste, and how is this cost distributed across different components of the computing infrastructure?
- Recent green computing research makes assumptions about the power draw and utilization of computing devices, often based on isolated research lab measurements: which of these assumptions hold in a larger enterprise setting and which do not?
- We heavily instrumented our infrastructure because we did not know what we would find; now that we have an understanding of the data, how would one design a measurement infrastructure to achieve good accuracy with the least effort?

The answers to these questions form the fundamental contributions of this paper:

- Detailed examination of where energy goes reveals that over 50% of the electricity is spent on computing. PC's account for 17% of the bill despite the fact that their utilization is very low. Networking equipment comes at 3.5% and shows no temporal changes despite variations in traffic load.
- Data analysis shows that estimating savings based on a few isolated desktop measurements is prone to errors due to the wide spread of PC power draws. Assuming that a day of power is representative and using it to calculate yearly values can be off by as much as 20%.
- Our deployment and data studies expose the relative importance of device coverage versus duration of deployment. Once a deployment is past the first month of data collection, one must prioritize the 'what to measure' question over the time scale of the study.

The rest of this paper reviews the current state of green computing data before diving into the analysis of the Powernet datasets. Along the way, it confirms or refutes a number of anecdotal observations, stressing the need for empirical data. The paper closes with guidelines for the design of future energy characterization studies.

## II. Background

Up until recently, the green computing community has had to rely on limited energy datasets, requiring researchers to make various explicit and implicit assumptions about the energy behavior of computing systems.

This section discusses some of the different ways in which related work has procured, used, and analyzed power data in the context of evaluating systems' research. At the end of the section, we formulate four common assumptions made in the context of green computing research.

In measuring only a small number of devices and generalizing claims, there is an implicit assumption that instances of the same equipment model or size have the same power characteristics. For example, in [16], researchers measure an individual desktop and LCD monitor and determine the ratio of power use between the two. Then, they collect aggregate data from a power strip that has another desktop and LCD of the same size. Applying the pre-computed ratio on the aggregate data leads to an estimate of the power draw of the second set of devices. Even though both desktops might have been Dells and both monitors 24-inch ones, we do not know how much variation there is between the seemingly identical devices.

A modeling approach that takes system subcomponents into consideration was used in [20]. Instead of collecting measurement with a meter, the authors use hardware components power models and software counters to calculate the power draw of a PC. This methodology was able to predict the power use of one machine based on a different one with 20% accuracy, indicating that even more sophisticated techniques that take device subcomponents into consideration will show error in estimation when assuming that similar equipment has similar power or usage behavior.

Prior works [8], [13], [16] have often based their analyses on power data collected from a small number of desktop measurements, 10 or fewer. In some cases, as shown in the introduction, evaluations are based on one or two values only. PCs are usually chosen based on convenience, i.e. desktops in the lab where the research is done, instead of using more deliberate samples.

Related work does not always indicate the duration over which power data was collected. In some cases [9], [16] power draw is shown over the course of a day or week, but the rest of the time it is presented as a one-time, instantaneous measurement. Despite the limited duration of measurements, they are used in calculations of long-term benefits of energy-saving techniques. Thus, in section VI we explore the accuracy of using a day or month of data to make full-year energy estimates.

The Energy Star program [3] establishes standards for energy efficient consumer products, including computing

| Device Type | Count |
|---|---|
| Desktop | 75 |
| Monitor | 70 |
| Laptop | 28 |
| Network Switch | 27 |
| Printer | 15 |
| Server | 36 |
| Thin Clients | 12 |
| Misc | 3 |
| **Total:** | **266** |

Table I
POWERNET COVERS A VARIETY OF DEVICES WHOSE POWER MEASUREMENTS ENABLE A CHARACTERIZATION OF THE ENERGY CONSUMPTION OF THE WHOLE BUILDING. SOME DEVICES ALSO HAVE CPU UTILIZATION OR NETWORK TRAFFIC MONITORS.

| Sensing Type | Num. datapoints |
|---|---|
| Power Data | 10 billion |
| CPU percent | 400 million |
| User Processes | 2 billion |
| Network Traffic | 10 million |
| **Total:** | |

Table II
SUMMARY OF COLLECTED DATA, ORGANIZED BY TYPE OF MEASUREMENT.

systems. As part of the process of obtaining an Energy Star certification, manufacturers of PCs submit power data for each of their devices. The result is an extensive database [6] of computer models and their power draw.

The lack of substantial power datasets within our research community means that it is not uncommon for academic works [12], [19] to base their analyses on Energy Star data. These data, however, are not representative of the real-world power characteristics of machines.

The observations in this section lead us to extract four assumptions:

- Devices of the same model or with the same specifications have low variation in power draw;
- Sampling a few devices in a class provides an accurate average measurement;
- Short-term measurements accurately reflect long-term power draw;
- Energy Star data is representative.

The rest of the paper analyzes a large power dataset and uses it to evaluate these assumptions. It also provids methodology guidlines for the future collection of energy measurements.

## III. DATA COLLECTION

The initial requirement for the Powernet deployment was the ability to sense individual computing devices. Commercially-available Watts Up .NET meters [5] were easy to obtain and became the first power sensor to be deployed. These meters were a useful first step in gathering power data, but had many practical issues. The closed nature of the firmware meant that we had no control over bugs and undesirable behavior (e.g., DNS lookup storms.)

The issues of scaling to hundreds of devices (Ethernet port availability and cable mess), high cost, and proprietary software hindered further deployment. By now many of the meters have either failed or have been upgraded. About 30 nodes remain active, in a sparse deployment around the building.

To scale the power-monitoring deployment, we designed custom wireless plug-level sensors. The sensing portion of these meters includes current and voltage sensors, plus a digital power meter chip that gives an instantaneous power reading [1]. The communications portion includes a low-power processor, a radio (2.4 Ghz, 802.15.4-based), and an integrated antenna. The meter software, built on TinyOS [18], includes sampling, routing [4] and dissemination [15] capabilities. The top-level application reads power draw every second and sends a data packet after buffering ten samples. The meters collect data via an ad-hoc multihop network using the Collection Tree Protocol [4]. The wireless power meters cost about \$110 apiece.

The current board , similar to the ACme [16], has been extensively tested and calibrated. We use a WattsUp meter in line with our power meters to calibrate them at different points between 0 and 300-watt loads. We find that raw meter values exhibit linear behavior with an r-squared of 0.99 or above for all meters. Of course, our calibration is limited to the accuracy of the WattsUp meters. Complementary to our work, [14] have designed a calibration method that can achieve utility-grade accuracy.

Instrumenting the entire Gates building was not feasible due to the costs and practical challenges associated with monitoring over 2000 devices. Yet, we wanted detailed-enough data to understand where in the building energy is spent and wasted. Several consideration went into deciding what to instrument. We focused our efforts on one of the two building wings, considering it representative of both wings in terms of types of devices and usage cases. Further, we were only interested in computing equipment, therefore we did not include miscellaneous electric loads such as staplers, fridges, coffee makers. or lights and HVAC. This is in contrast to the @Scale deployment [10], [14] which adopted a stratified sampling approach in order to avoid a random sample overwhelmed by small, insignificant loads. The main goal of Powernet's samples was to measure a wide variety of equipment to maximize the new information we gain. We did partially follow the stratified approach in allocating meters to go to different device categories such as servers and networking equipment.

Powernet takes a unique perspective on green computing by measuring not only device power draw but also device usage. We deploy a number of software sensor that collect utilization statistics in the form of CPU, active processes,

| Type/Count | Observed in 'whois' | Estimated by scaling | Total |
|---|---|---|---|
| Laptops | 47 | 29 | **76** |
| Low-end PC | 43 | 27 | **70** |
| High-end PC | 366 | 230 | **596** |
| **Total** | **456** | **286** | **742** |

Table IV
PERSONAL COMPUTERS ARE BINNED INTO THREE CATEGORIES, AND UNIVERSITY DATABASES AND ACTIVE NETWORK NODE COUNTS ALLOW US TO EXTRAPOLATE TO THE WHOLE BUILDING.

| Size | Count | Avg. Power |
|---|---|---|
| < 20" | 42 | 30 W |
| 20" to 22" | 40 | 45 W |
| 23" to 25" | 84 | 63 W |
| 26" to 27" | 15 | 80 W |
| 29" to 32" | 44 | 120 W |

Table V
A SURVEY SHOWS THAT MAJORITY OF BUILDING OCCUPANTS USE MID-SIZED LCD DISPLAYS. THE NUMBER OF LARGE (30") MONITORS IS INCREASING AS EQUIPMENT IS UPGRADED.

and network traffic. This utilization data is key for determining energy waste – the cases in which power is drawn but no useful work is done.

Tables I and II summarizes the different types of devices that our sensing infrastructure measures as well as the total number of datapoints for different sensing categories.

## IV. ENERGY CONSUMPTION

This section tackles the problem of extrapolating individual Powernet measurements to the whole building. We divide computing devices into four classes: LCD screen, PCs, networking equipment, and servers. Then, we combine power data with network activity logs, device registration databases, and a survey of building occupants to extend our observation to an energy picture of the full IT system.

We find that computing systems draw between 210 and 259 kilowatts, depending on the time of day, or 47% to 58% of the building's 445 kilowatt load. This aggregate power draw translates to 170,000 kilowatt-hours, or 50% of the building's monthly electricity usage.

Table III and Figure 1 summarize our extrapolation methodology and resulting breakdown. Ground truth is provided by aggregate measurements from outside the building, logged every 15 minutes by campus services. The top curve in Figure 1 shows one week of this data.

### A. Personal Computers

Personal computers are the second largest contributor to the energy consumption of computing systems, after servers. According to the department's database of registered devices there are about 1250 machines in the building that are actively observed on the network. Of those, roughly 740 are PCs.

In order to extrapolate to the whole building we bin PCs in three classes – laptops, low-end desktops, and high-end desktops. Low-end PCs are those with average power of 80 watts or less and include machines such a Mac Minis, Shuttle PCs, Dell Optiplex. Full-size desktops like the Dell Precision are considered high-end machines.

Next, we take the 742 desktop MAC addresses from the network database and cross-referenced them with the university's *whois* service. The whois metadata includes node descriptions, such as PC model and OS, provided upon network registration. Of the 742 nodes, 456 had description that allowed us to classify them as laptops, low- or high-end desktops.

Table IV shows the number of machines in each PC class; nodes with available description are labeled as 'observed' and breakdown of the other 286 assumes that the observed distribution is representative of the building. While there is no good way of verifying this assumption, it is a straightforward way of filling the gaps in inventory information.

Based on Powernet measurements, the median power draw for laptops is 26 watts, for low-end machines – 63 watts, and for high-end machines – 121 watts. This means that the three categories of machines draw 2 kW, 4.4 kW, and 72.1 kW respectively for a total of 78.5 kW a day or 58500 kilowatt-hours a month. The 742 personal computers in the building account for about 17% of total electricity consumption of the Gates building.

### B. Computer Displays

Powernet's measurements allow us to quantify the average power draw of different size LCDs; extrapolating to whole-building power draw requires the distribution of display sizes. To obtain an estimate of this distribution, we use an online survey asking occupants for the number, size, and manufacturer of the computer screens they use. Table V presents data from the 169 responses reporting 225 monitors. These responses account for 28% of the building's occupants. The table also shows the power consumption of different displays.

The cumulative power draw of the LCDs reported by users is 15 kW. Scaling that to the whole building yields a power draw of 52 kW or 12% of the building's power demand during daytime. Active duty cycling of screens reduces the energy footprint. Powernet data over time shows that displays are powered on 50 to 60 hours a week. Therefore, over one month, LCD screens consume about 14,000 kWh, or 4.2% of the monthly electricity budget.

### C. Server Machines

Powernet monitors 32 of the 500 servers in Gates Hall. Similar to desktops, servers exhibit varied power profiles. For example, a standard 1U rackmount can have a power draw anywhere between 95 and 275 watts. Unlike desktops,

| Device Type | Measured | Total | Extrapolated via | Total Draw | Uptime | Monthly Energy | Building Share |
|---|---|---|---|---|---|---|---|
| Switches | 27 | 62 | network admin records | 15 kW | 24 hrs/day | 11,000 kWh | 3.5% |
| Desktops/Laptops | 83 | 742 | whois, MAC registrations | 80 kW | 24 hrs/day | 61,000 kWh | 17% |
| LCD Displays | 70 | 750 | occupant survey | 48 kW | ≈8 hrs/day | 14,400 kWh | 4% |
| Servers | 32 | 500 | manual inspection | 117 kW | 24 hrs/day | 86,000 kWh | 26% |

Table III
WE CROSS-CORRELATE POWERNET MEASUREMENTS WITH IT DATABASES TO EXTRAPOLATE THE ENERGY CONSUMPTION OF ALL COMPUTING SYSTEMS IN THE BUILDING.

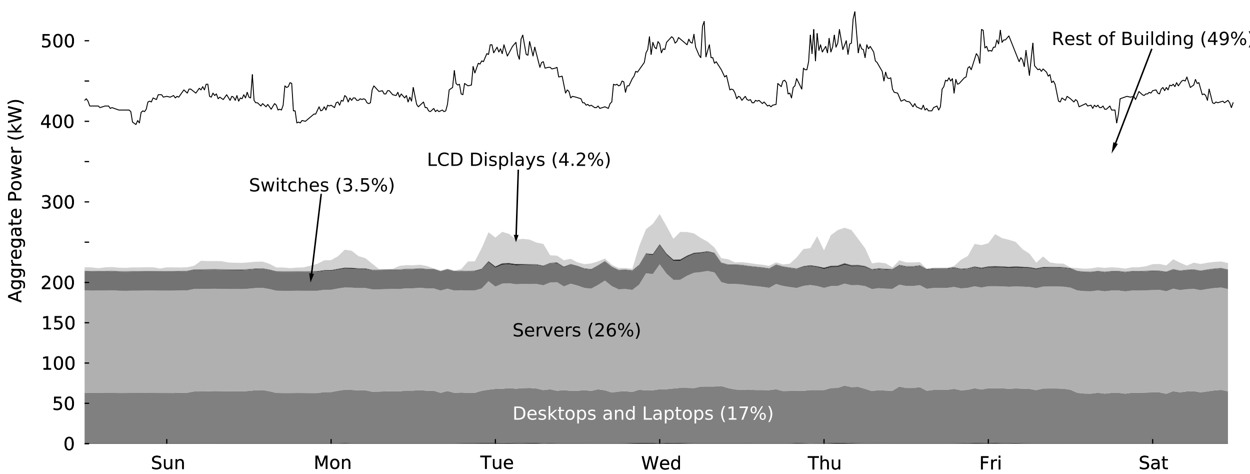


Figure 1. Aggregate power draw for the entire Powernet building shows diurnal and weekday/weekend patterns. Computing systems account for 51% of the total 445 kW. The given week of data is representative of the building, except for Monday, which was a university holiday (Feb 15).

the server population is much more homogeneous, e.g. 40 identical 1U machines in a single rack. Therefore, we spread out our measurements to get maximum coverage, with meters measuring identical devices for verification purposes. The average power draw we calculated from the Powernet measurements is 233 watts. With about 500 servers, the aggregate draw is 117 kW – 26% of the total building energy consumption per month. In the future, we hope extend the set of server measurement in order to have a more precise extrapolation.

### D. Networking Equipment

The network backbone in the Gates building is provided by 2 core switches located in the basement and 26 edge switches spread across the five floors. There are also a number of medium- and small-sized switches that have been deployed on as-needed basis. We account for all major switches and estimate the number of smaller ones with the help IT staff. Table VI summarizes the types of networking equipment together with their power draw. The power draw of wireless access points is folded into the switch data since they are Ethernet-powered.

Current networking hardware has constant power draw per linecard of active ports, with small power variation due to CPU load [17]. This observation together with the relatively small number and homogeneity of devices, lends well to a whole-building calculation. We use Powernet's measurements and inventory from Table VI to calculate the daily power draw of all networking equipment, 15.4 kilowatts. This translates to 11,500 kWh per month or 3.5% of the building's total consumption.

### E. Summary

This section presented a new methodology for characterizing power data, using both plug-level empirical measurements and device metadata, to create a detailed picture of IT energy. We find that 50% of the building's energy goes to computing equipment: 26% goes to servers, 17% to PCs, 4% to displays, and 3.5% to networking.

Our data confirms prior observations and intuition that PCs and servers are major contributors to the energy bill of enterprise buildings [14]. The data also highlights that smaller parts of IT, such as networks and LCD monitors, account for almost 8% of the overall building's electricity use. We find that displays are responsible for 50% of the building's diurnal power draw variation and are the only computing component that exhibits such patterns. This confirms that there is room for improvement not only in the IT infrastructure but also in the rest of the building.

## V. UTILIZATION

While a breakdown of the electric bill is a useful first step toward finding opportunities for savings, it is difficult to identify specific failures in energy efficiency. Energy

| Type | # Count | Power Draw (watts) |
|---|---|---|
| HP 5406zl (6-slot) | 20 | 325 |
| HP 5412zl (12-slot) | 8 | 500 |
| HP 2724 | 2 | 100 |
| Cisco Cat 6509 | 2 | 400 |
| Cisco Cat 4000 | 2 | 600 |
| Cisco Cat 3750G | 2 | 160 |
| Linksys | 2 | 50 |
| NEC (various) | 5 | 100 |
| Cisco (various) | 5 | 100 |
| Quanta (4-slot) | 5 | 50 |
| Misc (estimated) | 100 | 10 |
| **Total major switches:** | **53** | |

Table VI
SUMMARY OF SWITCH TYPES, QUANTITIES, AND AND ESTIMATED INDIVIDUAL POWER CONSUMPTIONS. THIS INVENTORY INCLUDES ALL MAJOR NETWORK SWITCHES AND EXCLUDES SMALL PER-ROOM SWITCHES AND HUBS.

| | Percentile CPU | | |
|---|---|---|---|
| **Machine Type** | $5^{th}$ | $50^{th}$ | $95^{th}$ |
| **Student PCs** | | | |
| Dell Precision T3400 | 0% | 1% | 7% |
| Dell Inspiron 530 | 1% | 1% | 8% |
| Dell Precision T3400 | 0% | 1% | 13% |
| HP Pavilion Elite m9250f | 0% | 0% | 25% |
| Dell Precision T3400 | 0% | 4% | 29% |
| High-end custom-built | 0% | 1% | 57% |
| Dell Optiplex 745 | 1% | 9% | 58% |
| **Staff PCs** | | | |
| Dell Dimension 9200 | 0% | 0.75% | 3% |
| Dell Precision 690 | 0% | 0.7% | 4% |
| Dell OptiPlex 760 | 0% | 0% | 5.45% |
| Dell OptiPlex SX 280 | 0% | 0.75% | 5.5% |
| Dell Dimension 9200 | 0% | 1.5% | 8% |
| Dell OptiPlex 745 | 0% | 1.5% | 9% |
| Dell OptiPlex SX 280 | 0% | 0% | 10% |
| Dell OptiPlex 760 | 0% | 1.55% | 17% |

Table VII
CPU UTILIZATION OF BOTH STUDENT AND ADMINISTRATIVE STAFF MACHINES REVEALS THAT PROCESSING RESOURCES ARE ONLY LIGHTLY TAXED. DATA WAS COLLECTED ONCE A SECOND FOR 11 MONTHS (STUDENTS) AND 1 MONTH (STAFF).

data alone is not enough, it is only meaningful if paired with a characterization of systems' utilization. This section examines the workloads of computers and network switches to determine what part of the energy is spent well and how much is wasted.

### A. Computers

Related work [7], [20] suggests that desktop machines are rarely turned off when not in use, and Powernet power measurements over a >1 year-long period support this claim. So far, green computing research has focused on solving the problem of idle PCs. Our utilization data sheds light on an equally wasteful problem – power-hungry machine that even when active, barely tax their resources.

Powernet collects data from both student and staff PCs and since the computing needs of the two groups are likely to differ, we consider them separately. Table VII shows the CPU utilization of a number of desktops. Computer science students use more of their available processing resources, but even so, in many cases CPU usage is under 30% for 95% of the time. The demand on administrative staff machines is even lower. Since most of the measured computers were left powered on at all times, the $50^{th}$-percentile data is not surprising: machines are often idling. What is surprising is that even when PCs are in use, the level of usage is low. If desktops were power-proportional that would not be an issue, but the current high baseline power draw means that the energy cost for a PC that is running at 5-6% of its capabilities is disproportionately high. In one extreme case, measurements showed that the most power-hungry staff desktop (quad-core Dell Dimensions 9200), drawing over 150 watts, has the lowest CPU utilization – 3.1% for 95% of the time.

Another way of investigating whether utilization matches the type of equipment we buy is to look at typical tasks users perform. We focus on staff computing because it is more representative of an enterprise computing environment. Table VIII shows the most common workloads on administrative machines, excluding Windows services and virus checks. The percentage of active time is calculated as the cumulative time over one month that the process was running; the range of time captures the minimum and maximum numbers over four computers. The workload data raises the question of mismatched user needs and technology. There is no reason why an entry level laptop or a Mac Mini cannot perform the same basic tasks (document editing, web browsing, PDF viewing) as a quad-core, 150-watt desktop.

Characterizing the utilization of computers has revealed that there is a lot more waste than idle machines alone. The baseline power draw of desktops, combined with low use of system resources, means that there are energy-saving opportunities even when PCs are actively used. Powernet's PC utilization data suggests that future green computing research should tackle all PCs, not just idle ones.

### B. Network Equipment

Section IV found that the networking infrastructure consumes 3.5% of the building's electricity monthly electricity. This translates to a cost of $15,000 a year just for networking. We also noted that switches consume a constant amount of power due to their hardware design. If the network is operating near capacity, then the 3.5% is energy spent well. Otherwise, if we find that the network operates at, say, 10% capacity even at peak, it means energy is wasted.

This prompts the questions of how much traffic is flowing through the 60 or so switches in the building, and whether smaller or fewer switches could more efficiently meet bandwidth demands.

| Process | % of time active |
|---|---|
| Acrobat Professional | 1% to 4% |
| Firefox | 0.5% to 4% |
| Internet Explorer | 0.3% to 2% |
| MS Excel | 1% to 2% |
| Thunderbird | 0.4% to 1.2% |
| MS Word | 0.2% to 0.8% |
| Outlook | 0.4% |
| Acrobat Reader | 0.3% |
| Explorer | 0.01% to 0.3% |

Table VIII
THE MOST POPULAR WORKLOADS ON ADMINISTRATIVE COMPUTING SYSTEMS ARE GENERAL OFFICE AND WEB APPLICATION. THESE WORKLOADS IMPLY THAT A LAPTOP CAN BE A USED INSTEAD OF A DESKTOP.

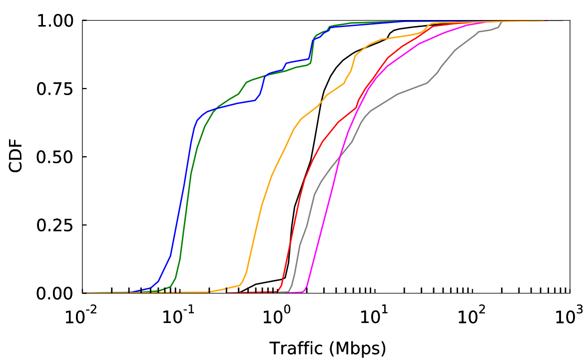


Figure 2. CDF of traffic for seven switches over 6 months shows that switches are operating well under capacity.

We begin by examining the traffic coming into one of the four switches on the second floor of our building. This is an HP Procurve switch with 96 1-gigabit active ports, consuming 500 watts and serving 50+ people. Over one week in March, bandwidth demand never exceeded 200 Mbps – an amount that could have been handled by consolidating it to a single gigabit port and providing 8-port switches (2 to 5 watts each) to users in individual offices.

To verify that this is not aberrant behavior, Figure 2 shows the cumulative distribution of traffic for 7 building switches. Note that the x-axis has a log scale. The number of ports for different switches varies from 24 to 120 and the CDF data was collected over 40 to 420-day periods.

Similarly to PCs, switches are highly underutilized. For the equipment we measure, total network demand is lower than 1000 Mbps 100% of the time. Network over-provisioning is not a new concept or observation; it provides benefits, including higher throughput, lower loss, and lower jitter. But when the average utilization is under one hundredth of one percent, several questions are worth considering. Is the amount of over-provisioning unnecessarily large? How can we take better advantage of the large amount of bandwidth that today's networks are ready to support? Going forward, there are two ways to address the issue:

| Device | #1 | #2 | % diff |
|---|---|---|---|
| Optiplex 760 | 60W (9) | 34W (24) | 43% |
| Optiplex SX280 | 68W (12) | 56W (8) | 18% |
| Optiplex GX620 | 71W (8) | 63W (13) | 11% |
| Precision T3400 | 117W (17) | 110W (10) | 6% |
| HP 5400zl switch | 467W (8) | 463W (4) | 0.01% |

Table IX
AVERAGE POWER DRAW OF TWO DIFFERENT DEVICES WITH THE SAME MODEL (STANDARD DEVIATION SHOWN IN PARENTHESES). TWO DEVICES OF THE SAME MODEL CAN DIFFER BY AS MUCH AS 43%. NETWORKING EQUIPMENT IS MORE UNIFORM THAN PCS.

consolidate equipment and make better purchasing decisions in the future, or make use of the extra available bandwidth.

The story that network traffic tells is no different than that of PC utilization – systems are heavily over-provisioned, often with no regard of expected workloads, leading to large energy wastes. Powernet's contribution is in bringing such utilization data to light and placing it in the context of green computing.

## VI. ASSUMPTIONS

The previous two sections revealed details about the energy waste in an office computing infrastructure, while Section II described some assumptions that have been made in the past, due to lack of rich energy-efficiency datasets. This section examines the validity of these assumptions in the context of Powernet's data. The hope is that future green research will benefit from data and methodology studies like this one and others [14].

### A. Assumption 1: Devices of the same model or with the same specifications have low variation in power draw.

Prior work has implicitly assumed that instances of the same equipment model or specification have the same power characteristics, simply because of the lack of better data. Under this assumption, measurements taken from one or two devices have been used to reason about other, unmonitored pieces of equipment. Unfortunately, such methodology can yield inaccurate results.

Powernet data reveals that some types of computing systems can exhibit large variations even when comparing two instances of the same device model. Table IX shows five example devices – 4 Dell desktop models and 1 network switch. The two Dell Optiplex 760 desktops have over 40% difference in their average power draw. In contrast, the two HP switches have almost identical power draw, as well as very low standard deviation over time. In some cases, even though the PCs appear to be the same on the surface, they might have been upgraded with custom components, causing a difference in power draw. Furthermore, while two devices from the same type might have similar motherboards, power supplies, and processors, they can differ in in the user workloads they support, leading to different power profiles.

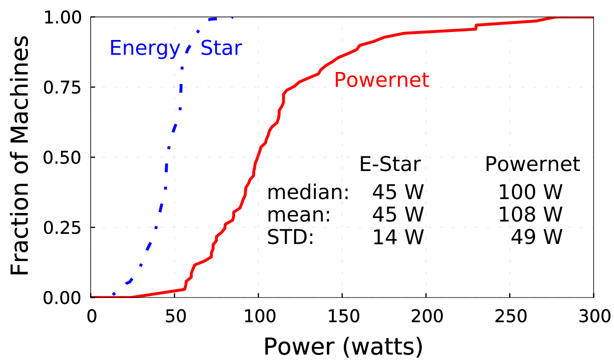


Figure 3. Energy Star data is not representative of real-world PC power.

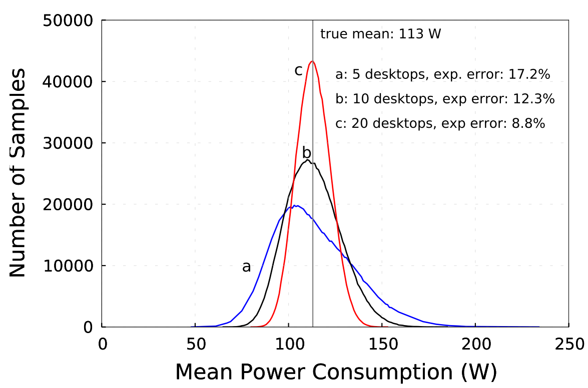


Figure 4. Desktop diversity requires the measurement of a large sample of the population. In this experiment, if only 5 desktops are used to estimate the power of all 69, then the expected error is over 16%.

There is no single solution to error in estimating power, when only partial measurements are available In the case of desktops, it is not surprising that there is great variability, but putting concrete numbers to it can help anticipate inaccuracies. Additionally, one could augment power measurements with other data, such as PC utilization, to get more accurate understanding of how equipment is used. In the cases when variability is low (e.g. switches), data points from only one or two devices can be treated as much more reliable.

### *B. Assumption 2: Sampling a few devices in a class provides an accurate average measurement.*

Figure 3, the Powernet curve, shows the wide distribution of desktop power. It is worth considering what errors can be expect if one were to sample only part of the PC population.

We use ground truth data from 69 desktops to show how the expected error of average power draw changes if based on random samples from the population. The average power draw of the 69 desktops is 109 watts. We generate 1,000,000 random samples of size 5, 10, and 20, drawing from the lists of 69 machines. Figure 4 shows the resulting histograms of estimated average power. Samples of only 5 desktops can have more than 16% error in estimating the mean power draw. Increasing the sample size from 5 to 20 machines cuts the error by more than a half.

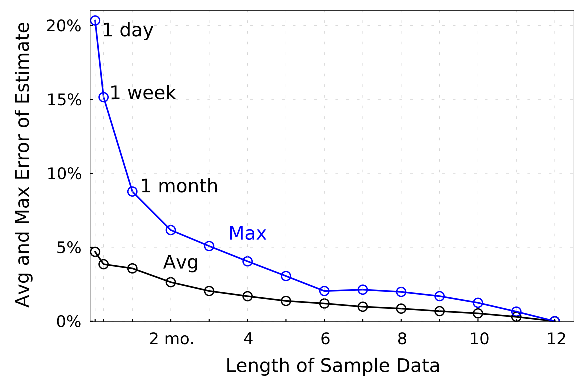


Figure 5. As the number of months of data increases, the standard deviation of error in estimates decreases. Even if only one month of data is used over 16 desktop, the year approximation will be within 4% of the true value of $1600.

The lesson is that when it comes to PCs, a small sample is not desirable if trying to extrapolate to a large, heterogeneous set. Recently work [10] correctly brings attention to the importance of complete device inventories in order to understand how varied an environment is and targeting measurement points accordingly.

### *C. Assumption 3: Short-term measurements accurately reflect long-term power draw*

The Powernet datasets show that while the base load of computing systems is consistently high, month-to-month variations do exist. These changes result in slightly different energy use throughout the month and year. For example, over one year the monthly power draw average of one desktop varied from 183W (min) to 293 (max), with 216-watt average over the whole year. Another PC was consistently between 247 and 257 watts. One question to tackle is ‘How does the duration of power measurements affect a yearly cost estimate for some set of devices?’

Our analysis uses data from 16 desktops; each PC was monitored for one year, from May 2010 to April 2011. The cumulative average power draw of the sixteen PCs is of 1524 watts ($1600 for the whole year, at $0.11 per kWh.) Examining the monthly average power draw of each machine reveals that no single month is representative of the whole year. If we were to take one day or week or month data in the hopes of estimating the yearly electricity cost, we should expect to be off. But by how much?

Using a single day of data from the year-long trace allows us to generate over 350 different estimates for the yearly cost. Similarly, for one week or month of data, we can compute multiple cost estimates. We can also repeat the process using sliding windows with size of two to 11 months. Figure 5 summarizes the results. The x-axis shows what duration of data was used for the estimate. The y-axis shows the average and maximum error of the estimate as percent of the real energy cost. In a worst case scenario, measurements taken for less than a week can have error of 15% or more. At the scale of a building IT systems, such error in predicting costs can be thousands of dollars. On the positive side, the analysis shows that collecting data at the month timescale, as opposed to longer, could yield data with an acceptable error. These results are in line with concurrent work at Lawrence Berkeley National Labs [11] who found that two months of data yields an acceptable tradeoff between deployment effort and accuracy.

### *D. Assumption 4: Energy Star data is representative.*

For a long time Energy Star was the only large openly available computer power dataset and while this is changing, it is worth to discuss why Energy Star data should be used cautiously. It is composed entirely of devices that have passed minimum energy efficiency requirements. It does not reflect the distribution of devices sold and data is self-reported. Furthermore, Energy Star measurements and certification do not consider PCs under load – they only deal with idle, sleep, and off states.

Figure 3 illustrates the divide between Energy Star data and the real-world measurements collected by Powernet. The differences are striking – close to 100% of the 4,000+ Energy Star desktops fall below the 100-watt cutoff. In our measurements, that is the median PC power draw. The Energy Star dataset has the benefit of a lot more data points and real-world distributions might shift from building to building. Using Energy Star in lieu of real measurements is likely to underestimate energy costs in most contexts but can be extremely useful in advocating lower-power machines.

## VII. Measuring Computing Power Accurately

The prior section pointed out four common assumptions in measuring computing power which can lead to inaccurate results. The results from Powernet, however, represent only one point in time. As computing continues to evolve, green computing research will need to periodically re-measure energy consumption and waste. This raises the follow-up question: ‘Given limited time, money, and effort, how should one measure computing system energy consumption in order to minimize error? ’ This section presents methodology considerations to guide future green computing research.

### *A. Step 1: Characterization*

Not all device classes are equal: some require much more effort to measure accurately than others. Table IX, for example, showed a 43% variation in the power draw of Optiplex 760 PCs but a 0.01% variation in the power draw of HP 5400zl switches. An approximate ordering of the different devices in terms of variability places desktops as the most diverse, followed by servers, laptops, LCD monitors, and lastly, switches. Rather than distribute measurement points uniformly, one should measure the high variation device classes more densely. But device classes change quickly: Dell, for example, no longer sells Optiplex 760 PCs. Being able to determine which device classes have significant variation requires up-to-date, current measurements.

To understand where to measure, one first needs to know which device classes are high variation and which are not. This can be done quickly, as a series of point measurements made over a day. For example, suppose that an enterprise has a large number of a new Dell PC. One can randomly select 10 of these PCs and measure each of them booting. This will provide a large dynamic range of power measurements within the class as well as across the class. If the 10 show significant differences, then they might need to be measured densely. One can use the observed power draw distributions and statistically compute what deployment of sensors will lead to the lowest observed error.

These point measurements should use simple digital readouts (e.g., Watt’s Up or Kill-A-Watt meters) which a person reads and writes down. Depending on a wireless mesh or wired network ports is probably more trouble than it’s worth (lack of connectivity, VLANs, etc.).

### *B. Step 2: Measurement*

Once a short-term characterization study has provided guidance on where to deploy sensors, they need to be deployed for a sufficient duration. We were able to use custom sensors and our own software to collect data over a wireless mesh, but this technology is not commonly available. Our experiences with Watt’s Up meters – coordination with IT infrastructure, reconfiguration, failure etc. – was that they are a poor choice for a very large, long-term deployment, but are acceptable at smaller scale.

The results in Figure 5 showed that energy consumption, especially for personal computing, changes significantly over time. One should measure for at least a week, and preferably for a month. After a month, expected error, even for high-variation devices, drops to 4%.

Given the choice between breadth (number of devices measured) and depth (length of measurement), greater breadth generally leads to more accurate results. At the extremes, it is better to gain a single point measurement of every device than measure one device for a year.

One last knob to consider is the interval at which measurements are taken. Wireless deployments often have limitations to the amount of data the network can handle. Concurrent work [11] found that 10-second intervals are a reasonable choice, capturing power dynamics without overloading the infrastructure. Our experiences showed that depending on the task, different resolutions are desirable. For many practical uses – visualizing data, computing long-term estimates – even 5-minute averages are useful. Higher-resolution data is needed for correlating utilization metrics with power draw.

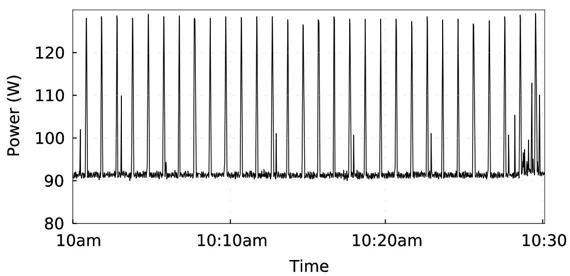


Figure 6. Power data collected once a second reveals a misbehaving PC. Earlier, 5-minute averages hid the anomaly. In certain use cases it is beneficial to have high-resolution data.

In several cases, we found that Powernet's 1-second data can pinpoint misbehaving equipment. For example, Figure 6 shows the raw data trace for one desktop. Initial analysis was using 5-minute averages and power draw appeared normal. A second look of the original data uncovered an aberrant behavior – a 30-watt spike once a minute – later discovered to be due to a bug in the Linux wireless drivers.

While different deployments will have varying resources and goals, experiences with Powernet teach us that having high-resolution data can be valuable not only for energy characterization but also for monitoring for unexpected behavior.

### C. Step 3: Extrapolation

The final step is to take the set of biased measurements and extrapolate to whole system power. Our experiences with Powernet have highlighted the need for data beyond power and utilization measurements. If extrapolation is to be successful, one also needs metadata in the form of equipment inventories and descriptions. Surprisingly, such metadata is not nearly as complete and readily available as we had hoped. Rather, we had to resort to indirect sources such as cross-correlating networked device registrations with active IPs on the network.

In the future, green computing researchers should encourage IT personnel to keep updated and detailed records of what equipment is added to a building.

## VIII. Conclusion

Characterizing the energy use of enterprise computing systems is the first step toward identifying opportunities for improvement. Extensive, empirical data allow researchers to better quantify the problems they are tackling and the potential impact of their proposed solutions. Powernet has provided such data and has shed light on some of the assumptions that we make when faced with the lack of solid measurements.

Despite our best attempts to cover as many computing systems for as long as possible, the Powernet data remain but a single study. While the exact breakdown of energy use and waste might shift from building to building, the overarching methodology and data analysis lessons remain. Going forward, green computing research has not only a reference dataset to use but also a blueprint for how to characterize enterprise building power given limited time and resources.

## Acknowledgments


We would like to thank the anonymous reviewers and our shepherd, Randy Katz, for their invaluable comments and fresh point of view. We also thank Wanja Hoffer, Mayank Jain, and Jung Il Choi for the insightful discussions and suggestions. This work is supported by the Department of Energy ARPA-E under award number DE-AR0000018. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.


## References


[1] ADE7753: Single Phase Multifunction Energy Metering IC with di/dt Input. http://www.analogdevices.com.

[2] Department of Energy, Annual Energy Review, October 2011. http://www.eia.doe.gov/aer/.

[3] Energy Star. http://www.energystar.gov.

[4] TEP 123: Collection Tree Protocol. http://www.tinyos.net/tinyos-2.x/doc/.

[5] Watt's Up Internet Enabled Power Meters. https://www.wattsupmeters.com/secure/products.php, 2009.

[6] Energy Star Computer Power Data. http://www.energystar.gov/index.cfm?c=revisions.computer_spec, 2010.

[7] Y. Agarwal, S. Hodges, R. Chandra, J. Scott, P. Bahl, and R. Gupta. Somniloquy: Augmenting Network Interfaces to Reduce PC Energy Usage. In *Proc. Networked Systems Design and Implementation*, 2009.

[8] Y. Agarwal, S. Savage, and R. Gupta. SleepServer: A Software-Only Approach for Reducing the Energy Consumption of PCs within Enterprise Environments. *USENIX Annual Technical Conference*, 2010.

[9] Y. Agarwal, T. Weng, and R. Gupta. The Energy Dashboard: Improving the Visibility of Energy Consumption at a Campus-Wide Scale. *ACM Workshop On Embedded Sensing Systems For Energy-Efficiency In Buildings*, 2009.

[10] R. Brown, S. Lanzisera, H. Y. I. Cheung, J. Lai, X. Jiang, S. Dawson-Haggerty, J. Taneja, J. Ortiz, and D. Culler. Using Wireless Power Meteres to Measure Energy Use of Miscellaneous and Electric Devices in Buildings. *Energy Efficiency in Domestic Appliances and Lighting*, 2011.

[11] H. I. Cheung, S. Lanzisera, J. Lai, R. Brown, S. Dawson-Haggerty, J. Taneja, and D. Culler. Detailed Energy Data Collection for Miscellaneous and Electronic Loads in a Commercial Office Building. *To appear in ACEEE Summer Study on Energy Efficiency in Buildings*, August, 2012.

[12] K. Christensen, C. Gunaratne, B. Nordman, and A. George. The Next Frontier for Communications Networks: Power Management. *Computer Communications*, 2004.

[13] T. Das, P. Padaala, V. Padmanabhan, R. Ramjee, and K. Shin. LiteGreen: Saving Energy in Networked Desktops Using Virtualization. *USENIX Annual Technical Conference*, 2009.

[14] S. Dawson-Haggerty, S. Lanzisera, J. Taneja, R. Brown, and D. Culler. @scale: Insights from a Large, Long-Lived Appliance Energy WSN. *11th ACM/IEEE Conference on Information Processing in Sensor Networks, SPOTS Track (IPSN/SPOTS '12)*, 2012.

[15] J. W. Hui and D. Culler. The Dynamic Behavior of a Data Dissemination Protocol for Network Programming at Scale. In *Proc. ACM Conference on Embedded Networked Sensor Systems (SenSys)*, 2004.

[16] X. Jiang, S. Dawson-Haggerty, P. Dutta, and D. Culler. Design and Implementation of a High-Fidelity AC Metering Network. In *Proc. Conference on Information Processing in Sensor Networks (IPSN)*, 2009.

[17] M. Kazandjieva, B. Heller, P. Levis, and C. Kozyrakis. Energy Dumpster Diving. In *Workshop on Power Aware Computing and Systems*, 2009.

[18] P. Levis and D. G. et al. T2: A Second Generation OS For Embedded Sensor Networks. Technical Report TKN-05-007, Telecommunication Networks Group, TU Berlin, 2005.

[19] S. Nedevschi, S. Ratnasamy, J. Chandrashekar, B. Nordman, and N. Taft. Skilled in the Art of Being Idle: Reducing Energy Waste in Networked Systems. In *Proc. Networked Systems Design and Implementation*, 2009.

[20] J. Reich, M. Goraczko, A. Kansal, and J. Padhye. Sleepless in Seattle No Longer. *USENIX Annual Technical Conference*, 2010.