The Megawatts behind Your Megabytes: Going from Data-Center to Desktop

David Costenaro and Anthony Duer, EnerNOC Utility Solutions

ABSTRACT

Businesses and individuals are migrating in droves to cloud-based software, offsite data warehouses, and streaming media services; clamoring to shed the responsibilities of owning and maintaining IT equipment and infrastructure. But should energy issues drive us to be MORE conscious of this world – of what goes on behind our flatscreens when we surf 23 websites in a tabbed browser or stream a whole season of "The Sopranos" in the background while we cook?

Research on computational energy use has occurred in many segregated clusters, from end use devices to data centers to servers, but what happens at a system level? What happens between the server rooms and our screens? Has there been disproportionate focus on "data-supply" efficiency; and should demand-side management techniques be applied to promote "data-use-efficiency"?

This paper is a thought-piece on the how's and why's of end-to-end, IT energy use. It will pursue questions like: What type of equipment is used to get a MB from the data-center to your desktop? Is multi-tabbed browsing the IT equivalent of leaving the refrigerator door open? How much energy does it use? How much does it cost; and who pays for it?

Our major finding is that the Internet uses an average of about 5 kWh to support the utilization of every GB of data, which equates to about $0.51 of energy costs. Only 38% of those costs are borne by the end-user, while the remaining costs are thinly spread over the global Internet through which the data travels; in switches, routers, signal repeaters, servers, and data centers (See Figure 1 below). This creates a societal "tragedy of the commons," where end users have little incentive to consider the other 62% of costs and associated resources.

Figure 1. Internet Energy Breakdown (kWh per GB)

Introduction

We'd like to begin by describing a vital national industry. The U.S. Department of Homeland Security has deemed it a critical infrastructure sector and a multitude of stakeholders at the national, regional, and state levels carefully monitor and regulate its comings and goings. It is an industry that has a history of dynamic and transformative growth, driven by exciting technological advancements and ever-mounting consumer demand. Indeed, consumers have begun to expect the services offered by this industry as a given, almost as a right. At a moment's notice, it is expected to be ubiquitous, reliable…and cheap.

Such expectations of pervasiveness and affordability have tipped an industry migration to larger companies, larger capital structures, and larger investments in more centralized and sophisticated production centers to drive economies of scale. This focus on highly-efficient central processing and production has yielded considerable gains, but the pace of improvement is slowing as low-hanging fruit is exhausted and theoretical or economical limits are approached.

Now this industry is on the cusp of a new revolution; what will it do to meet growing consumer demand as resources become scarce and paths forward become mired with constraints. Peak capacity and congestion issues abound. Difficult "last mile" problems creep up; where large delivery arteries must step down into the aging, crowded infrastructure of tiny local capillary networks. Some providers hope to leave the central production model, sacrificing economies of scale to situate closer to the consumers. An air of uncertainty lingers as new startups and established giants alike seek to forge the industry's future…

What industry are we describing…?

Similarities between the Energy and Information Technology Industries

The truth is, the paragraphs above could have been written about either the Energy or Information Technology (IT) industry. These two industries have a startling number of parallels, and would do well to learn from each other. Indeed, the two industries are acutely dependent on the services of each other; and must go beyond simply observing each other from across the fence. They will increasingly need to integrate and co-mingle their services and strategies to understand and embrace their unique interdependence. Internet companies will need to develop staff such as energy managers, energy czars, and sustainability directors. Utilities will need to stop viewing IT departments as cost centers and welcome them, along with smart grid and network experts, into the C-suite[1] where real decisions and planning are taking place. Figure 1 below shows the parallel form and relative pervasiveness of the IT and energy industries.

[1] The group of officers in an organization with the word "chief" in their titles; e.g., chief executive officer, chief financial officer, or chief operations officer.

Figure 2. Similarities: Internet Topography and Power Grid Infrastructure in the U.S.

The Role of Demand Side Management

The particular parallel between these two industries that we would like to explore in this paper is that of Demand Side Management (DSM). As mentioned above, the dominant industry focus to date has been to drive more and more production out of centralized plants – electric power plants in the case of electric grids and data centers in the case of IT networks. In the past, consumer behavior was treated as a given, and you designed the equipment around it.

Only in recent decades has the energy world shifted from this perspective. There is now a sharpening focus on end-to-end efficiency; for transmission, distribution, and perhaps most notably, the end-use consumer. This is being driven by escalating energy prices, peak capacity constraints, last-mile problems, and legislation regarding emissions, renewable generation, and energy efficiency. This recent focus on the demand side of this industry, as opposed to the supply side, is fundamentally changing the way consumers view, use, and think about energy.

It is our view that the Information Technology industry is not far behind with respect to this DSM paradigm. IT research is advancing rapidly, keeping technology costs down and access to bandwidth at least acceptable, although the U.S. lags other developed countries in this area (FCC 2010). Many people are familiar with Moore's law, which states that computational speeds are increasing at an exponential pace (Wikipedia 2012). There is also a corollary to this relationship known as Koomey's law, which states that computational energy efficiency is also increasing at an exponential rate (Koomey 2009). Is this a license, however, to do as much computation as we want? To use the Internet without regard? It is true that per unit computational power and unit energy efficiencies are improving, but it is also true that the absolute energy usage of computers and the Internet has risen dramatically and is projected to continue rising in most future scenarios (EPA 2007).

Distribution and throughput demands are rising exponentially, creating many of the same problems and issues that are challenging the energy industry. Much debate and planning is ongoing across the country regarding how the Internet should or shouldn't be regulated and how customers should interact with it. The authors have noted a lack of information and corresponding lack of urgency around the demand side of the IT industry, and hope that this paper can help in the simple but critical act of informing relevant parties. Further, we hope this

paper will precipitate discussions around ways that consumers can (wittingly or unwittingly) reach beyond their screens and touchpads to be active participants in data-use efficiency and end-to-end system solutions.

Internet Infrastructure

In order to understand the energy impacts of data-use, we first need to understand the underlying equipment. The Internet is a global system of interconnected computer networks that uses a very specific set of protocols to organize itself. It is a network of networks that consists of millions of private, public, academic, business, and government networks that are linked by a broad array of electronic, wireless, and optical networking technologies (Wikipedia 2012). Communication is enabled between any of these nodes by means of sending standardized data "packets." The sending device breaks the data into packets and sends each one bouncing independently from one node to the next along a dynamically optimized route of carrier nodes. Each packet may take a different route and arrive in a different order, but they are reassembled correctly at the receiving node according to the protocol's specification, thus giving the Internet its amazing robustness. The infrastructure is generally set up as shown in Figure 3 below, with three tiers of successively more advanced points of presence (POPs) to transmit packets between end-use devices.

Figure 3. The Infrastructure of the Internet

An example transmission activity might begin on a desktop computer when an end user requests to download a song. That request gets converted into packets that are sent to their wireless router and then through a fiber optic line to a local internet service provider (ISP), which is a Tier 3 POP. The packets know where to go because they first consult a domain name server (DNS), which is like a giant phone book to find where the requested song file is stored on the Internet. Once the ISP processes the packet, it sends it along, up to a Tier 2 POP through a fiber optic cable with more bandwidth. At this level of bandwidth and above, special routing equipment[2] is used to layer multiple signals on top of each other in the same cable strand. This is called wavelength-division multiplexing (WDM) and can overlay more than a hundred distinct signals in the same strand, all of which are electronically filtered and deciphered by the routing equipment. The packet with the request might then be passed up to a Tier 1 POP and shuttled across the country before hopping back down to a lower tier POP to find the data center where the song is stored.

All of the sends and receives at the carrier nodes are brokered by routers, switches, or hubs; each humming with the electronics of their own processers and overhead loads such as cooling, power conditioning, and lighting.[3] Additionally, when fiber optic, copper-wire, or wireless communication links must span long distances, the signal degrades and must be regenerated periodically by repeaters, each of which adds to the energy footprint of the activities. Once the signal finally arrives at the data source, the download begins, sending the song in packets back to the original requester by way of the same process in reverse.

Energy Footprint Analysis

As we have seen, energy is used by many different pieces of equipment in many different places to enable an Internet communication, but each POP and end user only has a small purview of the end-to-end energy use of the entire process. There are so many network activities running simultaneously that it is difficult to dissect and investigate any one download, transmission, email, etc. We attempt to explore these energy relationships by using both a top-down and a bottom-up analysis approach that meets in the middle to calibrate and check for reasonableness. This means first defining the total usage of both energy and data by the Internet (top down) and secondly, dividing it up into representative shares among the participating equipment (bottom up). The amount of energy and data associated with an internet activity is then determined by the equipment it engages at each POP it touches over whatever amount of time the activity takes. Network traces can be used to determine the path, number of POPs, and duration of any internet activity.[4]

We began our calculations of Internet power consumption at the total level (Raghavan 2011). Raghavan and Ma's estimate showed wall socket power of the internet between 84 and 143 GW. We then refined this estimate by subdividing the equipment and activities into the tiers described above. This included a different overhead power estimate for each tier, as well as the

[2] The state of the art for such equipment today is the reconfigurable optical add-drop multiplexer, or ROADM (Wikipedia 2012).

[3] Data Center Power Usage Effectiveness (PUE) is an important industry metric which is defined as the ratio of the total data center power consumption divided by the power consumption of the IT equipment. The average PUE for U.S. data centers is 1.91 (EPA 2009) while state-of-the art designs for large cloud service providers achieve power usage effectiveness levels as low as 1.1 to 1.2. (Microsoft 2009).

[4] For a listing of trace route services, see: http://www.traceroute.org. For an example trace route service, see: http://www.net.princeton.edu/traceroute.html.

addition of signal repeaters as a "transportation" element. These changes produced our updated estimate of Internet power consumption in Table 1 and Figure 4 below. The minimum power estimate is 103 GW and the maximum estimate is 180 GW, with a simple average of 141 GW.

Table 1. Internet Equipment and Energy Use

					Portion of Use Dedicated to Internet Activities				
Category	**Equipment**	**Count**	**Wall-Socket Power (W)**	**Wall-Socket Duty Cycle**	**Min Weighting**	**Max Weighting**	**Total Min Power (GW)**	**Total Max Power (GW)**	**Total Average Power (GW)**
End Use Device	Desktops	750,000,000	150	50%	0.50	0.95	**28.1**	**53.4**	**40.8**
End Use Device	Laptops	750,000,000	40	50%	0.75	1.00	**11.3**	**15.0**	**13.1**
End Use Device	Smartphones	1,000,000,000	1	50%	0.25	0.90	**0.1**	**0.5**	**0.3**
Trans3	Wi-Fi/LAN	100,000,000	20	100%	0.75	1.00	**1.5**	**2.0**	**1.8**
Trans3	Cell Towers	5,000,000	3,000	100%	0.10	0.50	**1.5**	**7.5**	**4.5**
Trans3	Telecom Switches	75,000	75,000	100%	0.00	0.25	**0.0**	**1.4**	**0.7**
Trans3	km of Fiber Optics	500,000,000					-	-	-
Trans3	Optical repeaters (per 75km fiber)	6,666,667	400	100%	0.50	0.90	**1.3**	**2.4**	**1.9**
Trans3	km of Copper	3,500,000,000					-	-	-
	Signal repeaters								

Trans3	(per 75km copper)	46,666,667	400	100%	0.10	0.50	**1.9**	**9.3**	**5.6**
POP Tier 3	Servers	40,000,000	300	100%	0.50	0.95	**6.0**	**11.4**	**8.7**
POP Tier 3	Routers	400,000	5,000	100%	0.90	1.00	**1.8**	**2.0**	**1.9**
POP Tier 3	Cloud	20,000,000	300	100%	0.80	1.00	**4.8**	**6.0**	**5.4**
POP Tier 3	Overhead power	100% *(PUE=2.0)*					**12.6**	**19.4**	**16.0**
Trans2	km of Fiber Optics	500,000,000					**-**	**-**	**-**
Trans2	Optical repeaters (per 75km fiber)	6,666,667	600	100%	0.50	0.90	**2.0**	**3.6**	**2.8**
POP Tier 2	Servers	30,000,000	300	100%	0.50	0.95	**4.5**	**8.6**	**6.5**
POP Tier 2	Routers	300,000	5,000	100%	0.90	1.00	**1.4**	**1.5**	**1.4**
POP Tier 2	Cloud	15,000,000	300	100%	0.80	1.00	**3.6**	**4.5**	**4.1**
POP Tier 2	Overhead power	75% *(PUE=1.75)*					**7.1**	**10.9**	**9.0**
Trans1	km of Fiber Optics	500,000,000					**-**	**-**	**-**
Trans1	Optical repeaters (per 150km fiber)	3,333,333	800	100%	0.50	0.90	**1.3**	**2.4**	**1.9**
POP Tier 1	Servers	30,000,000	300	100%	0.50	0.95	**4.5**	**8.6**	**6.5**
POP Tier 1	Routers	300,000	5,000	100%	0.90	1.00	**1.4**	**1.5**	**1.4**
POP Tier 1	Cloud	15,000,000	300	100%	0.80	1.00	**3.6**	**4.5**	**4.1**
POP Tier 1	Overhead power	25% *(PUE=1.25)*					**2.4**	**3.6**	**3.0**
GRAND TOTAL							**103**	**180**	**141**

Table 1 Notes:

- *Telecom switch power includes 50% overhead load for cooling (Raghavan 2011).*
- *Signal repeater estimates: Linear spacing (Wikipedia 2012); Power usage (Fujistsu 2012).*
- *Overhead power/ PUE estimates: Tier 1 (Microsoft 2009); Tiers 2 and 3 (EPA 2009).*
- *Please note that actual equipment specifications run along a broad and diverse spectrum, but we use average, representative values to facilitate a manageable and useful modeling exercise.*

Figure 4. Total Internet Energy Usage (GW)

If the average energy draw of the Internet worldwide is 141 GW as our analysis suggests, the system requires well over 100 base load power plants. Using the segmented analysis in Table 1, we estimate that only 38% of that load is associated with end user devices, 14% is associated with transmission and communication aspects, and 48% is devoted to network resources in data centers and server rooms.

CPU and Network Activities

The data above signify that internet activities have real, measurable energy consequences. The question for us as individual data users and a society as a whole is whether they warrant attention and action or they are too small to notice. What happens on a micro-scale when an individual end user is performing tasks over the Internet? In Figure 5 below, we used the "Resource Monitor" to observe a single laptop while performing some common tasks.[5] First we noted the baseline CPU and Network activity with no programs or websites active. Then we opened the top ten websites according to Alexa.com in a single tabbed browser and noticed that there was considerable CPU and network activity while opening and interacting with those websites (Alexa 2012).[6] Even when those 10 tabs were left alone for several minutes, periodic blips of system usage continued to occur, associated with data refreshing, downloading of banner ads, and other activities.

[5] Configuration as tested: Resource Monitor in Windows 7, Internet Explorer 8, running on Lenovo ThinkPad X220 laptop computer with EnerNOC Analyst system configuration. Data running on AT&T network in St Louis, MO.
[6] Top Ten websites as of 3/9/2012: Google, Face book, YouTube, Yahoo!, Baidu, Wikipedia, Windows Live, Blogspot, Twitter, and QQ.

Figure 5. CPU and Network Activity for Common Tasks

System requirements were then viewed for a single tab playing a song from a streaming music service and separately for a single tab showing a television show from a streaming video service. The streaming music showed slightly increased CPU usage relative to baseline, and very similar network activity except for periodic blips associated with data buffering and ad refreshes. The streaming video resulted in similarly increased CPU usage, but sustained and markedly higher network activity, with 3 to 4 Mb per second to support the larger data transfer associated with video.

The Energy Cost of Data

Data usage on the internet is estimated to be 20,151 PetaBytes per month (Cisco 2011). This is equivalent to 241 billion GB per year. Applying these figures to the average power estimate yields a figure of **5.12 kWh per GB**. Assuming an average electricity cost of $0.10-US per kWh (EIA 2012), this equates to a system energy cost of data at **$0.51 per GB**.

This means that streaming an average 3 GB movie costs $1.54 in energy, while an average 5 MB song can stream for energy costs of only $0.003. These costs are thinly spread among the various network nodes that it touches. Some examples for common Internet related tasks are shown below in Table 2.

Table 2. Energy Cost of Common Internet Activities

Energy Cost per Visit

Activity	Data Transacted	End User	Transport-ation	Data Center/POP	TOTAL	Source
Visit Amazon.com - Main Page	567 KB	$0.0001	$0.0000	$0.0001	$0.0003	Firefox - Page Info
Streaming 5MB MP3 song	5,120 KB	$0.0010	$0.0004	$0.0013	$0.0026	Estimation

Watching 5 Minute YouTube Video	13,313 KB	$0.0026	$0.0009	$0.0033	$0.0068	YouTube - Calculated from reported kbps
Streaming 3GB Movie	3 GB	$0.5869	$0.2098	$0.7415	$1.5381	Estimation
Playing The Orange Box (Online Video Game)	16 GB	$3.0322	$1.0838	$3.8311	$7.9471	Steam Online Gaming Platform

Improving Data Use Efficiency

Given all the above information, what options do we have for containing the energy footprint of the Internet? There are already many strategies being rigorously researched, pursued, and implemented (DOE 2012, EPA 2012). Of course computational efficiency is increasing, per Koomey's Law as discussed above. Large data centers are making rapid advances in efficiencies and pulling customers to the cloud; driven by several business drivers like economies of scale, diversification of server load, flexibility, and the ability to sidestep organizational IT policies and hurdles (Koomey 2011). Additionally, manufacturers are shipping their products with more pervasive use of sleep modes and power saving modes as default selections.

Until recently, computers and network equipment were built to have no variability in power, but simply processed or sent zeros until a meaningful packet came along. Thus, for these systems, the energy used by the communications infrastructure would barely change, even if end users completely curtailed their data demand. This is changing, but a good portion of existing equipment remains like this, with no built-in energy intelligence or strategy. This may also serve as a caveat to the calculations given in this paper; if we present a figure of 5.12 kWh per GB, reduction of data use won't always necessarily lead to a proportional reduction of energy use. In areas of the Internet where local equipment will retain fixed energy use levels regardless of traffic, there may be no reduction. When aggregated, though, the concept of kWh/GB makes sense for the entire system. If an aggregate number of GB capacity needs to be served, the proportional amount of power will surely need to be there.

Implications for Everyday Internet Users

What does this mean for individuals? A number of suggestions are outlined below to increase our awareness and practice of data-use efficiency:

- **Raise awareness that streaming media have energy usage associated with them**. There is energy used by the Internet to deliver streaming music and video to your device. This is no different of course than traditional broadcast music and video, so keep in mind the energy conserving ethos that has always existed for these technologies: minimize

their on-time when no one is using them. Indeed, some streaming services will pause periodically if they are not interacted with – both as a way to avoid paying royalty fees to the contributing artists and as a way to reduce unnecessary energy use. Importantly, many data providers are beginning to curb their unlimited data plans, levying additional charges if large caps are exceeded. This is a first step toward sending a rational price signal to consumers.

- **Use tabbed browsing reasonably.** If tabs are open and the computer does not move into a sleep mode, data will continue to flow to the open tabs. Be aware of extended periods of time when it makes more sense to add a bookmark and come back later.
- **Limit unnecessary data transmission**. When you are sending emails or attachments, be mindful of whether it needs to be sent multiple times or to multiple people. Compress attachments if applicable. Consider turning off the feature that automatically includes the entire previous email thread on all email replies. Turn off "push notifications" and wifi searching that you do not need to conserve wireless battery life.
- **Activate powersave & monitor sleep modes**. A standard efficiency measure that has been available for years, these modes are often not the default option when shipped from the manufacturer, and need simply to be activated.
- **Use a power strip or smart power strip** to shut off peripherals & chargers when not in use. These measures are becoming more commercially available, affordable, and user-friendly.

Conclusions

We hope that we have shed some light on this subject for decision makers in both the energy and information technology fields. These industries will continue to have profound effects on each other.

We understand that we may have overlooked elements of this subject matter, and welcome any constructive comments on ways to improve the analysis and enhance the conversation.

We want to emphasize that we do not view internet use, or energy use for that matter, as intrinsically bad and something to be avoided. Both have been hugely instrumental in raising standards of living and solving problems the world over. They also have profound symbiotic effects on each other. In fact, a large portion of energy used within the Internet has enabled a new era of sensors, controllers, data systems, and software to perhaps save more energy in other end uses than the Internet uses itself (Lohr 2011). This does not mean, however, that we can turn a blind eye to the way that the Internet uses energy.

There seems to be a lack of urgency around the demand side of the IT industry, and we hope that this paper helps in the simple but critical act of informing relevant parties. Our findings show that the average end user of the Internet is only aware of and responsible for paying a small portion of the costs linked to their actions. This may prove unsustainable in the face of future growth trends for data demand. Individual computer energy use impacts transmission, communication, and network resources in a very real way. We need to start thinking about usage patterns, data paradigms, and the customer experience differently; learning from lessons that the energy industry has taught us about DSM.

In addition to Energy DSM, we need to start seriously considering Data DSM.

References

Alexa. 2012. *Top Sites*. San Francisco, CA: Alexa Internet, Inc. Retrieved March 9, 2012, from www.alexa.com

Cisco. 2011. *The Internet in an Exabyte World*. San Jose, CA: Cisco Systems, Inc. Retrieved March 9, 2012, from www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html

[DOE] US Department of Energy. 2012. *Data Center Projects and Resources*. Washington, D.C. http://www1.eere.energy.gov/manufacturing/datacenters/

[EIA] US Energy Information Administration. 2012. *Electricity Average Retail Price Data*. Washington, D.C. Retrieved May 4, 2012, from http://205.254.135.24/electricity/monthly/epm_table_grapher.cfm?t=epmt_5_3

[EPA] US Environmental Protection Agency, ENERGY STAR Program. 2009. *ENERGY STAR Data Center Infrastructure Rating Development Update*. Washington, D.C.

[EPA] US Environmental Protection Agency, ENERGY STAR Program. 2007. *Report to Congress on Server and Data Center Energy Efficiency*. Washington, D.C. www.energystar.gov/ia/partners/prod_development/downloads/EPA_Datacenter_Report_Congress_Final1.pdf

[EPA] US Environmental Protection Agency, ENERGY STAR Program. 2012. *Server Virtualization Resources*. Washington, D.C. http://www.energystar.gov/index.cfm?c=power_mgt.datacenter_efficiency_virtualization

[FCC] Federal Communications Commission. 2010. *The National Broadband Plan*. Washington, D.C. Retrieved May 4, 2012, from www.broadband.gov/plan/

Fujitsu. 2012. *FLASHWAVE Product Line Specification Sheets*. Retrieved May 4, 2012, from www.fujitsu.com/

Koomey, Jonathan G. 2011. *4 reasons why cloud computing is efficient*. GigaOM, San Francisco, CA. Retrieved March 9, 2012, from http://gigaom.com/cleantech/4-reasons-why-cloud-computing-is-efficient/

Koomey, Jonathan G, Stephen Berard, Marla Sanchez, Henry Wong. 2009. *Assessing Trends in the Electrical Efficiency of Computation Over Time*.

Lohr, Steve. 2011. *The Internet Gets Physical*. The New York Times, New York, NY. http://www.nytimes.com/2011/12/18/sunday-review/the-internet-gets-physical.html

Microsoft. 2009. *Microsoft's Top 10 Business Practices for Environmentally Sustainable Data Centers*. Seattle, WA.

Raghavan, Barath and Justin Ma. 2011. *The Energy and Emergy of the Internet*. ICSI and UC Berkeley.

Wikipedia. 2012. *Internet*. FL: Wikimedia Foundation, Inc. Retrieved March 9, 2012, from http://en.wikipedia.org/wiki/Internet

Wikipedia. 2012. *Moore's Law*. FL: Wikimedia Foundation, Inc. Retrieved March 9, 2012, from http://en.wikipedia.org/wiki/Moore's_law

Wikipedia. 2012. *ROADM*. FL: Wikimedia Foundation, Inc. Retrieved March 9, 2012, from http://en.wikipedia.org/wiki/ROADM

Wikipedia. 2012. *Fiber-Optic Communication*. FL: Wikimedia Foundation, Inc. Retrieved March 9, 2012, from http://en.wikipedia.org/wiki/Fiber-optic_communication