HAL
open science

ICT Energy Demand: what got us here won't get us there!

Rabih Bashroush, Eoin Woods, Adel Noureddine

To cite this version:

Rabih Bashroush, Eoin Woods, Adel Noureddine. ICT Energy Demand: what got us here won't get us there!. IEEE Software, 2016, 33 (2), pp.18 - 21. 10.1109/MS.2016.53 . hal-01897887

HAL Id: hal-01897887

https://hal.science/hal-01897887v1

Submitted on 17 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ICT Energy Demand: what got us here won't get us there!

Rabih Bashroush (University of East London), Eoin Woods (Endava) and Adel Noureddine (University of East London)

Background

IT activity has entered a new era. In 2008, the number of things connected to the internet exceeded the number of people (Evans, 2008), while Cisco predicted that in 2011, 20 typical households would generate more traffic than the entire internet in 2008 (Evans, 2008). We are witnessing an explosive growth in data driven by more affordable storage systems and the proliferation of mobile, IoT, social media, and smart cities to name a few. Add to this the software applications being created around this data, from business analytics to apps that can monitor your cow's pregnancy through a sensor connected to its tail and send you an SMS when she will most likely calf. The vast majority of the applications being developed are cloud-based, meaning constantly increasing demand on data centres, where many of the large Tech companies are continuously expanding their capacity.

Currently, data centres consume a substantial amount of energy and are thought to produce more Greenhouse Gas emissions than the entire aviation sector. In 2013, data centres in the U.S. alone consumed an estimated 91 billion KWh of electricity, and this is expected to rise to 140 billion KWh by 2020 (Delforge, 2014). It is not surprising therefore that a recent survey of data centre managers showed that power density and energy efficiency are among their top current and future concerns (Intel Corporation, 2012). This is particularly important for Colocation and Managed Service data centre providers who are based within or close to large cities where grid power availability is limited.

So, can data centres cope with this demand and continue to support the exponential growth of the big data and cloud revolution? We cannot be sure, but we can see that the challenges are widely recognised: a number of related areas are the subjects of active research, ranging from new cooling technologies to more energy efficient servers and building designs, to runtime workload consolidation and management techniques. However, the software architecture community has been slower to recognise their potential contribution and to mobilise to meet this challenge. Addressing energy efficiency at the architecture level is still far from being mainstream. Can we continue designing systems without any consideration of their energy and power efficiency and let others worry about running them in an energy efficient way? Should energy efficiency be a bolt-on system property, or a quality attribute that is addressed at the design stage?

Why aren't Software Architects helping, yet?

Software architects may not be prioritising energy efficiency for a number of reasons. Firstly, we currently have very little understanding of the impact of design decisions on energy efficiency or an understanding of how it affects other system qualities such as user experience, reliability and performance. Without this knowledge, it is difficult to perform trade-off analysis to understand the benefit or cost of improving energy efficiency. Minor changes to the system design could yield substantial benefits, such as avoiding unnecessary polling or eliminating redundant housekeeping tasks that prevent equipment from entering lower-power states. However, a lack of relevant design tools and frameworks mean that it is still difficult to achieve more sophisticated optimisations that include consideration of contextual information about the runtime environment.

Second, in order to achieve the next order of magnitude in energy efficiency, we need to think outside traditional design boundaries. This will require people from different specialisations and departments to work together. This can often be difficult to achieve, given current organisational software governance structures, with different teams sometimes having competing objectives, not to mention human dynamics and political barriers. Current technologies also provide few mechanisms to allow communication across different technology layers (that is, the application software, middleware, hardware, network, cooling, power infrastructure, etc.) which would enable cross-layer optimization.

Finally, so far, energy efficiency rarely features as an end-user requirement or concern. On one hand, there is the problem of split incentives where operators of systems (e.g. administrators or data centre managers) do not pick up the energy bill (which tend to come out from the facilities budget). Accordingly, they would see very little return from any savings made from energy efficiency. On the other hand, ICT energy costs, given current energy prices, do not constitute more than 1% to 3% of a typical organisation's budget. So, when efficiency is pursued, it is easier to achieve it by addressing areas with larger budget share (e.g. payroll!). This problem is exacerbated by the lack of benchmarks, metrics and reliable data to allow realistic comparisons of different energy efficiency opportunities and their realistic returns.

What do Software Architects Think?

In an attempt to understand the practicing architect's perspective on energy efficiency, we surveyed a small number of representative architects from organisations across a number of domains. We asked them whether they had to deal with an energy efficiency related challenge in a recent project (over the past five years), and whether they felt they were equipped with the right tools to address such challenges. Additionally, we asked architects whether they believed energy efficiency would be a major architectural concern over the coming five years. The results from the survey are discussed below.

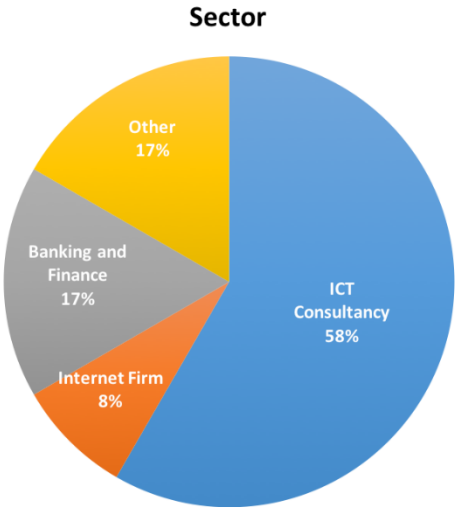

Figure 1. Breakdown of participants by sector

Figure 1 above shows the breakdown of participants based on the type of organization that they worked in. Although the survey was small-scale (confidence level of 80% and error margin of 20%), it can be seen from the pie chart that the respondents were representative of the IT sector based on organisation type.

So what did they say? As can be seen from Figure 2 below, while the majority of surveyed architects did not have to deal with energy efficiency concerns over the past five years (83%), most thought this would be a major concern over the next five years (67%). Yet only a quarter of them agreed or strongly agreed that they had the right tools to address these challenges (Figure 3).

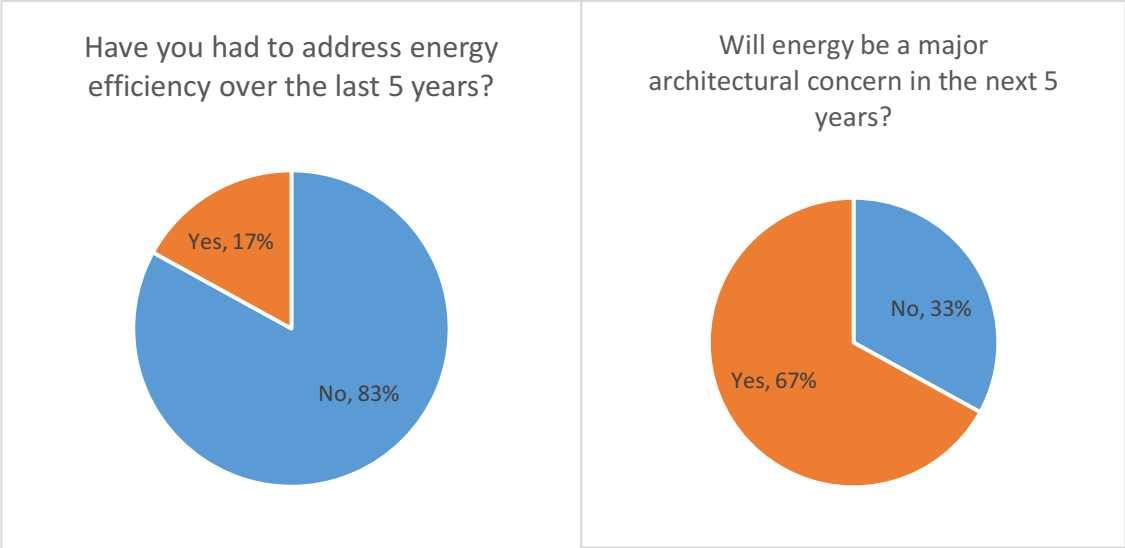

Figure 2. Breakdown of Prominence of Energy Efficiency

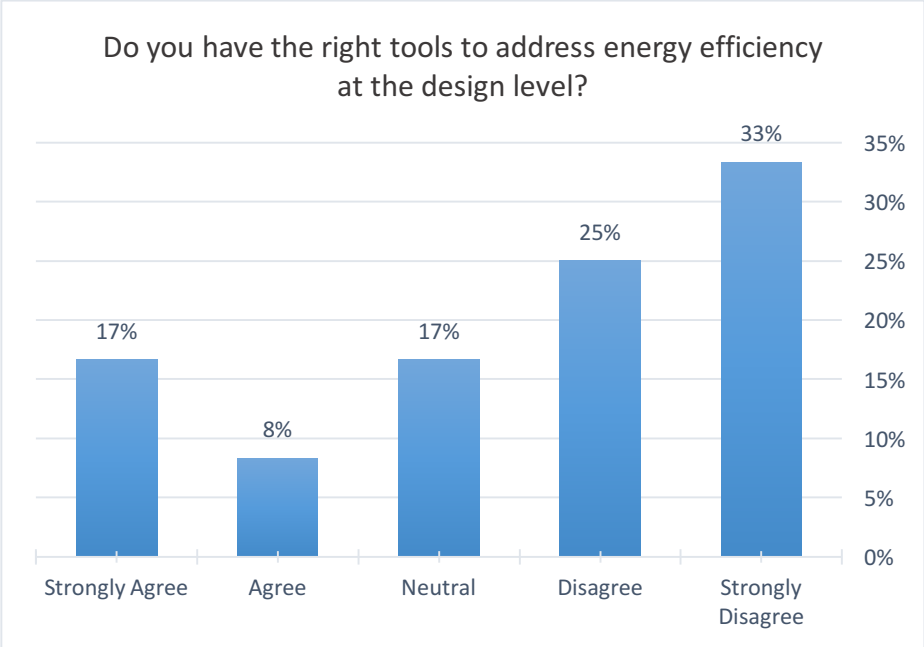

Figure 3. Do Architects have the right tools to address energy efficiency?

The survey results confirmed our view of the challenges that software architects face with respect to energy efficiency, particularly around the lack of tool support to address energy at an architectural level, and the current low priority that many stakeholders place on addressing energy concerns in solution design.

Opportunities & Future Directions

Although this appears to be a gloomy situation for energy efficiency, in fact a range of potential solutions and energy saving opportunities do exist, we just need to recognise them and learn to exploit them.

Starting from the deployment view of a system, when considering placement of workload across the data centre, we can base allocation on the cooling profile, which consumes on average about 40% of the total data centre energy, rather than just the power or performance profiles often used today.

Then when designing applications, we could move beyond today's typical SLA based system quality requirements. Most SLAs for throughput, availability and performance are rigid time-based measures that do not consider energy usage as a factor. If we moved towards specifying these requirements in more flexible ways, based on outcomes over time, this would allow many applications to include energy efficiency and pricing as part of their design. By taking into account the time of the day, the current real-time energy price and other environmental factors, some types of large applications can manage their processing in real-time to minimise their energy cost.

Another direction to consider as application designers is how we match different types of workload to different types of processing environment in order to minimise energy costs. Today we tend to focus on ease of application construction and the performance of different application platforms. If we can make estimates of energy consumption available for different types of platforms then this is another dimension to consider. For example, vector-based hardware architectures can offer considerable energy savings for some types of data intensive applications, while running some types of mainstream hardware at lower clock speeds reduces energy consumption while having little impact on the perceived performance for many types of application.

The other trend that we feel is inevitable is the eventual energy rating of software products. Just as industrial equipment became regulated so that its energy efficiency was stated clearly, and then consumer domestic appliances, today our IT hardware is energy rated but our software is not. Given the rising tide of environmental awareness, the likely ongoing rise in energy prices and the ever increasing IT workloads, it seems very likely that software too will become energy rated.

The metric we are missing in order to rate the energy efficiency of software is a measure of the work performed by a piece of software for a certain amount of energy consumed. This isn't something we can do easily today, but it is another active research area, with researchers investigating a range of approaches. The basic idea is to identify a representative characteristic operation for a particular piece of software (perhaps sending or receiving a million messages for a message bus, or processing 1000 single item orders via a web shop) and to measure the energy consumed to process that workload. This allows us to calculate the Application Work Energy Efficiency (AWEE) as the work_performed / energy_consumed (European Commission, 2014) and allows us to compare the energy efficiency of similar applications.

In summary, there is little today that we as software architects can simply take "off the shelf" and start using to monitor and minimise the energy consumption of our software. However, only a few years ago, this could also have been said of our colleagues working in

data centre and infrastructure architecture. In contrast, today, they need to be aware of and using standards such as the EU Code of Conduct on Data Centres (European Commission JRC Insitute for Energy and Transport, 2015) and Green Grid Data Centre Maturity Model (The Green Grid, 2015). A similar process for software seems inevitable and so now is the time for us as software architects to start understanding this emerging area and making sure that our contributions and concerns are heard.

References

Delforge, P. (2014). *America's Data Centers Consuming and Wasting Growing Amounts of Energy.* New York: Natural Resources Defense Council.

European Commission JRC Insitute for Energy and Transport. (2015). *The European Code of Conduct for Energy Efficiency in Data Centres.* European Commission. Retrieved from http://iet.jrc.ec.europa.eu/energyefficiency/ict-codes-conduct/data-centres-energy-efficiency

Evans, D. (2008). *The Internet of Things Infographic*. Retrieved October 23, 2015, from CISCO: http://blogs.cisco.com/diversity/the-internet-of-things-infographic

The Green Grid. (2015). *The Green Grid Data Centre Maturity Model.* Retrieved from http://www.thegreengrid.org/en/Global/Content/white-papers/DataCenterMaturityModel

Bios

Rabih Bashroush is a Faculty Member and Director of the Enterprise Computing Research Group at the University of East London. Contact him at: r.bashroush@qub.ac.uk

Eoin Woods is the chief technology officer at Endava, a European IT services company. Contact him at eoin.woods@endava.com.

Adel Noureddine is a researcher in software engineering and green computing at the University of East London. Contact him at: a.noureddine@uel.ac.uk