Is it really Greener in the Cloud? An investigation of energy trends in Cloud Computing.

Chris Browne, Haley Jones, Paul Compston

Research School of Engineering, Australian National University

Contact: Chris.Browne@anu.edu.au

Abstract

Cloud computing has become a popular method of IT deployment in the last decade, most visibly through the rise of online services and social media. Recent reports claim that businesses can achieve significant reductions in greenhouse gas emissions by shifting their processes to the cloud. These studies do not consider the change in user interaction with data when it is in the cloud, such as through mobile devices. This paper investigates this energy-saving claim by reviewing trends in and between four domains of this problem and identifying their key drivers in a business setting. These domains considered are on-site computing, cloud infrastructure, data transport, and device adoption. By mapping a causal loop diagram to explain how the dominant trends are linked between these domains, we conclude that overall energy consumption is likely increase when moving computing processes to the cloud.

1. Introduction

Cloud computing is the latest paradigm in computing[1] (Buyya, 2008), and is changing the way that people and businesses interact. It has also been marketed as a 'green' computing alternative for business, with claims that there are significant energy savings, and hence cost savings, by moving traditional infrastructure to a cloud-based computing model (CDP 2011, Google 2011).

This paper examines the energy-saving effect of cloud computing by examining trends in energy consumption. In order to formulate our dynamical hypothesis, major trends have been identified in the domains of *on-site computing, cloud infrastructure, data transport* and *device adoption*. Previous studies examine and account for energy consumption in these 'problem' domains (see Table 1), but do not consider a dynamic relationship between them. We propose that an approach that considers feedback between the problem domains may draw different conclusions from the previous work. A causal loop diagram is used in this paper to propose these feedback mechanisms.

The utility of computing technology has improved significantly through the use of cloud services, through new communication platforms such as those in social media, and through new collaboration applications. However, the validity of energy-saving claims needs to be discussed and debated. If cloud computing allows people to access more data through a greater number of platforms, then per capita energy consumption is likely to increase unless there are other mechanisms that reduce energy consumption. This becomes critical if a perception that the cloud is 'greener' leads to a misinformed adoption, where there is in fact higher overall energy demand.

In Section 2, we provide a brief background to the reports that prompted this paper. In Section 3 we look at previous work and the dominant trends in each of the four domains. In Section 4, a causal loop diagram has been mapped to examine the feedback structures between the four domains. In Section 5 we look at 'what if' scenarios, and conclude our study in Section 6 with our recommendations for a numerical model.

1. Cloud computing has many definitions, from full-blown application platforms to browsing the Internet. The definition of cloud computing we have used in this investigation considers the cloud as interaction with data as a unified resource that exists beyond the business's immediate systems. Similar definitions call this the public cloud.

2. Problem Articulation

In 2011, the Carbon Disclosure Project (CDP) released a report finding that a typical, large company transitioning their human resources system to the cloud could reduce their carbon emissions significantly[2] (CDP 2011).

This paper does not seek to question the specifics of the CDP paper and model. Our point of concern is to question its simple cause-and-effect message: *that cloud computing reduces energy consumption*. This message was picked up by many news sources and technology blogs, and was used by cloud service provider Google to show that switching to their cloud-based email service, Google Mail, was almost 80 times more efficient than running in-house email (Google 2011).

By conflating the utility benefits of cloud computing with apparent energy saving benefits, advocates of cloud computing may be misleading the public about the overall environmental benefits of adopting cloud services. This could lead to a paradoxical situation that, by moving business operations to the cloud to reduce energy consumption, the overall consumption actually grows.

To investigate the claim further, we will look at a typical business setting transitioning their services to the cloud. The intention is to match the CDP paper's scenario[3] by using similar context. The trends of each of the problem domains will be examined in the next section.

3. Problem Domains and Trends

The following domains were identified as areas to investigate the energy consumption of cloud computing:

- *On-site Computing:* energy consumed by the user's workstation and on-site servers
- *Cloud Infrastructure:* energy consumed by the user at the data centre[4]
- *Data Transport:* energy consumed by the transport of the user's data between the site and the data centre
- *Device Adoption:* energy consumed by new platforms available to access the data

2. The CDP report claimed that a typical Food and Beverage company could reduce CO2 emissions by 30,000 metric tons over five years in the cloud, when compared with a business as usual scenario. In an economy-wide scenario, US businesses with annual revenues over $1 billion could cut CO2 emissions by 85.7 million metric tonnes annually, when compared with a business as usual scenario.
3. The key aspects of the CDP scenario are: a Food and Beverage firm of $10 billion annual revenue, with 60;000 employees across 30 countries.
4. Data centres have been chosen as the focus for this domain as they are the back-end of cloud services.

Recent research focuses on single domains of cloud energy consumption, rather than the relationships and feedback between them. In this section we investigate the dominant trends in each domain, and in Section 4 we propose how each of the domains listed above interact through a causal loop diagram.

Table 1 maps how three recent reports consider energy consumption due to the cloud. A column for this paper is included in this table as a comparison.

Table 1: Comparison of considerations in key recent studies into overall cloud energy consumption. Considerations classed as (S)tatic, (D)ynamic or (N)ot Considered based on discussions in each paper.

Domains	**Carbon Disclosure Project (2011)**	**Google Mail (2011)**	**Baliga *et al* (2010)**	**This paper**
Scenario aspects of model mapped to domains identified in this paper				
On-site Computing	Workstations (S) Servers to zero (D)	No change on-site (N)	PC or laptop (S)	Workstations (D)
Cloud Infrastructure	Energy per server (S) Data centre efficiency (D)	Google Mail servers (S)	Server energy (S)	Data centre efficiency (D) Data centre demand (D)
Data Transport	Size of firm (S) Energy/transaction (S)	No change in data (N)	Range of scenarios (D)	Transport efficiency (S) Change in volume (D)
Device Adoption	None (N)	None (N)	None (N)	Change in device adoption (D)
General model information				
Report Focus	Cost of transitioning large business to cloud service	Traditional email compared to Google Mail service	Relative energy change with different data consumption patterns	Relationship between domains
Time period	Forecast 2011-2020	Point in time	Point in time	To be determined[5]
Scale	Large business extrapolated to market	Small, medium and large business	Single user/instance	Single user in large business
Energy Conclusions	Cloud more efficient than business as usual	Cloud email 80 times more efficient	Varies based on scenario	To be determined

Table 1 categorises the previous research into the domains considered in this paper. It shows that the previous research has not examined all the domains that we consider to be important when calculating the total energy consumption of the cloud. We have been unable to find data that comprehensively illustrates the total energy consumed by the cloud. With the data available to us, in this section we establish dominant trends for each of the domains, and identify likely drivers within each domain. In Section 4 we look at the relationships between these trends.

3.1. On-Site Computing

The first domain considered is on-site computing, which has been simplified in this scenario as only the energy consumed by user workstations.

5. This is not determined at this stage of the study, but it is expected that data will be sufficient to examine trends and propose a forecast for the period 2000-2025.

The energy consumed by computers within a business has been of concern for many decades. This is demonstrated most visibly in the 1992 introduction of the Energy Star program, which raises awareness of the energy efficiency of technology in a modern office environment. However, Bray (2006) shows there is little technical compliance with the program[6], and user non-compliance with power management practices, such as turning computers off in idle times, varies between organisations[7]. The effective implementation of a power management strategy across a business could reduce energy consumption considerably. The other driver we found for energy consumption is the increase in computing utility, which will now be explained.

We have described the main driver of workstation energy consumption as *computing utility*, which can be explained through two trends: improvements in computing performance, and increasing workstation screen size. To demonstrate this trend, we have compiled data on the maximum power draw[8], screen size and release dates of Apple Inc.'s series of all-in-one computers, which have been in production since 1984 (Apple 2012). The reason for using an all-in-one machine is that it represents a consistent overall form factor, rather than a highly variable energy consumption of a multitude of desktop and monitor configurations.

---

6. The Energy Star programs largely require computers to meet Idle, Sleep and Standby settings. However, there is low industry compliance, with only 21% of 141 surveyed desktops meeting all the requirements (EPA, 2006). An Australian study showed that only 64% of a surveyed 22 computers advertised as meeting Energy Star requirements actually met them (DEHWA 2009). In addition, the 'In Use' energy efficiency on Desktop computers relies on the energy efficiency of the power pack, rather than a compliance to an amount of energy consumed.

7. Webber *et al* (2006) observed that 60% of workstations were left on overnight. The results, however, are quite varied, with Bray (2006) observing that anywhere between 0%-91% of computers were turned off out of business hours.

8. Koomey *et al* (2009) notes that using nameplate power to draw conclusions on power consumption is erroneous, with measured energy consumption generally significantly lower. This may be the case in these data in Figure 1; however, the trend would still suggest that energy consumption is increasing over time.

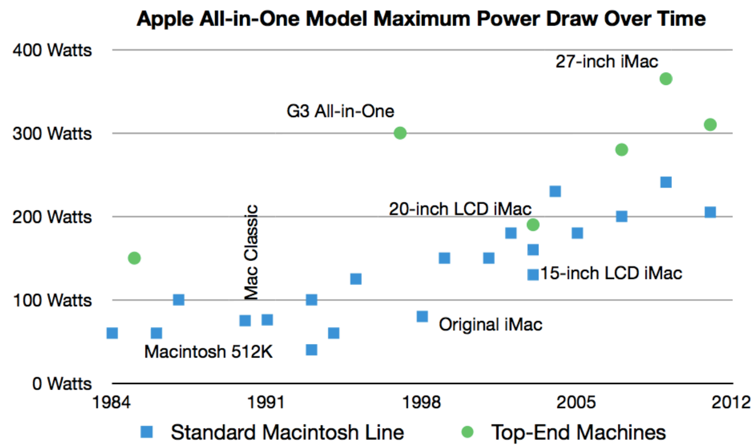

Figure 1: Apple all-in-one model maximum power draw over time (Compiled from Apple 2011). The trend shows that maximum power draw is increasing over time.

The trend in power draw over time in Figure 1 is somewhat counter-intuitive, especially as this product line has stated in its specifications that it meets Energy Star guidelines since 2002. However, the Energy Star rating only considers idle and standby power draw, which explains why there is no incentive to keep active power draw low.

We propose that this increased maximum draw is due to an increase in computing utility, which we have identified as a continuous increase in processing performance (see Moore's Law, Moore 1965) and increasing screen size (Figure 2). The trend in increasing energy consumption with larger screens is apparent, even though there was a significant reported energy saving when monitors shifted from CRT[9] to LCD[10] technology[11].

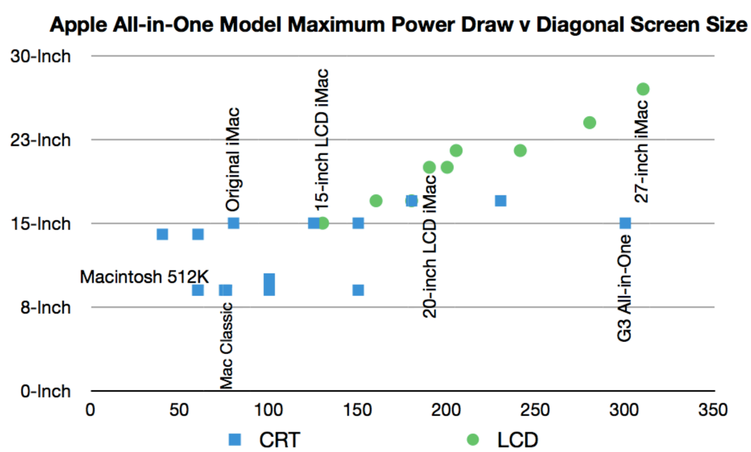

Figure 2: Apple all-in-one model maximum power draw v diagonal screen size (Compiled from Apple 2012). The trend shows all-in-one energy consumption increases with screen size.

9. cathode ray tube
10. liquid crystal display
11. A typical LCD screen draws one-half to one-third of the power of an equivalent-sized CRT monitor (IEA 2009)

The combination of low adoption of power management strategies and the trends in Figure 1 and Figure 2, leads us to the conclusion that as computers are replaced through business cycles in the workplace, the overall energy consumption is likely to increase. This is shown through an approximate four-fold increase in energy consumption between the years 1984-2012 (see Figure 1), and has been included as a driving influence for this reason.

Domain Drivers (On-site Computing)

In summary, the domain drivers for on-site computing energy consumption have been identified as:

- pressure to improve energy efficiency leads to increased power management compliance, which reduces on-site energy consumption
- pressure to improve workstation utility leads to investment in technology, which increases energy consumed.

We also considered areas that may be of concern but for which available trend data was not able to suitably address:

- user patterns on power management compliance, and whether this changes over time
- whether old workstations or servers are actually retired from the system, or new applications are found for them

In the following section we investigate how the cloud infrastructure affects the overall energy consumption.

3.2. Cloud Infrastructure

The cloud infrastructure in this study has been defined as the energy consumed by the data centre, as data centres are the major back-end to cloud services. In Koomey's (2008) definition of energy consumption in data centres, the network that connects the data centre to the user is not included[12]. For this reason, we consider data transport as a separate domain. However, Koomey's (2008) definition does include the infrastructure in the data centre, such as heating and cooling, and the internal network.

A breakdown of energy consumption trends for worldwide data centres shows that the total energy required to power data centres more than doubled between 2000 and 2006 (EPA 2007). A more recent estimate suggests that this general trend has continued, though

12. We include data transport energy in Section 3.3.

it is thought that an increase in adoption of virtualisation — where multiple virtual machines can run on a single physical server — and the recent economic downturn stalled growth in the global installed base. These data are shown in Table 2.

Table 2: Energy consumption of data centres worldwide[13] (from data in Koomey 2008). 2010 Lower and Upper represent the best estimates to date (Koomey 2011).

	Units	2000	2005	2010 (lower estimate)	2010 (upper estimate)
Total Energy	BkWh[14]	70.8	152.4	203.5	271.8
Volume servers	%	27.8%	33.1%	33.0%	30.0%
Mid-range servers	%	9.5%	4.4%	2.5%	2.7%
High-end servers	%	4.0%	2.8%	3.8%	4.8%
Storage	%	4.0%	4.9%	9.4%	8.9%
Communications	%	4.8%	4.8%	6.0%	5.7%
Infrastructure	%	50.0%	50.0%	45.2%	47.9%
Calculated PUE	ratio	2.0	2.0	1.82	1.91

These data suggest that there is a move away from mid-range servers, and an increase in the need for storage capacity. The Power Usage Effectiveness (PUE) measures efficiency through a ratio that compares the total energy, and the energy consumed by computing. This metric is used to compare the efficiency of data centres.

Significant improvements can be made in the efficiency of data centres, with major cloud providers reporting their PUEs below 1.5 (Katz 2009). However, data in Table 2 suggests that there have only been modest improvements across the board on global data centre PUE.

Domain Drivers (Cloud Infrastructure)

In summary, the domain drivers for cloud infrastructure energy consumption have been identified as:

- an increased demand for cloud computing directly leads to increased off-site energy consumption
- increased demand also provides opportunity for data centres to improve their efficiency through technologies that allow a data centre to achieve a lower PUE.

In the following section, we investigate the dominant trends in the energy required to transport data between the user and the data centre.

---

13. PUE has been calculated from these data using the formula: PUE = Total Energy / Computing Energy. Columns should total 100%, but may not due to rounding.
14. Billion kilowatt hours

3.3. Data Transport

The third domain we examined was data transport, which is a significant contributor to off-site energy consumption. This was discounted in the Google (2011) analysis, which states:

> *We would expect network energy to increase somewhat, as more traffic must traverse the Internet in the cloud-based solution. However, this effect is secondary to the large effect on server energy.*

Baliga *et al* (2010) investigates the trends in data transport in an office setting. They conclude that energy can be a large factor for energy consumption in cloud computing, and describe scenarios where the cloud does and does not provide opportunities for savings.

Using their example of the public cloud as a storage service, they show that transport consumes the vast majority of energy, in a dynamic relationship with storage and the servers, when almost any amount of data is transferred. This relationship is shown in Figure 3a. Figure 3b compares the energy required access a 1.25MB file on a laptop hard drive or through the cloud. Their analysis shows that the total energy consumed is less through the cloud, unless the download rate exceeds approximately 5MB/hour.

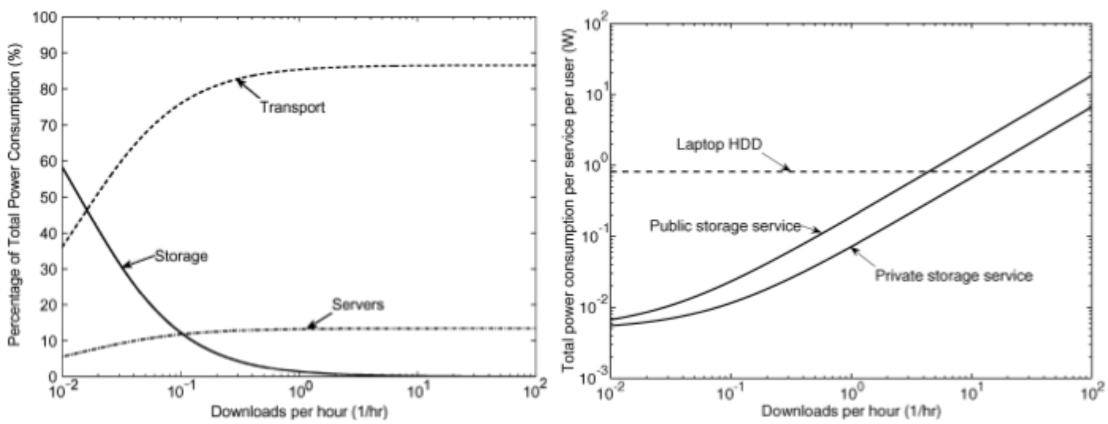

Figure 3: a) Left: Relative percentages of total power consumption in the public cloud; b) Right: Total power consumption per service per user from (Baliga et al 2010). In this example, the file downloaded is 1.25 MB. Transport makes up a significant percentage of total energy consumed when data is downloaded[15].

The analysis in Figure 3 (Baliga *et al* 2010) does not take into account the energy required by the end user's device, only the transport of data. Baliga *et al* (2010) acknowledge that accessing data from the cloud with a laptop would consume the energy

15. Private cloud storage shown in Figure 3b consumes less energy in transport, as it passes through fewer switches, routers and exchanges.

to power the server, transport the data, and run the device. This observations highlights the justification behind the three domains discussed thus far. It does not include the device adoption, which will be discussed in the following section.

A second major trend in data transport is that the volume of data transported is increasing over time. Data averaged from a Nielson (2011) report in US smartphone data usage is displayed in Figure 4. This displays a trend that consumption is increasing.

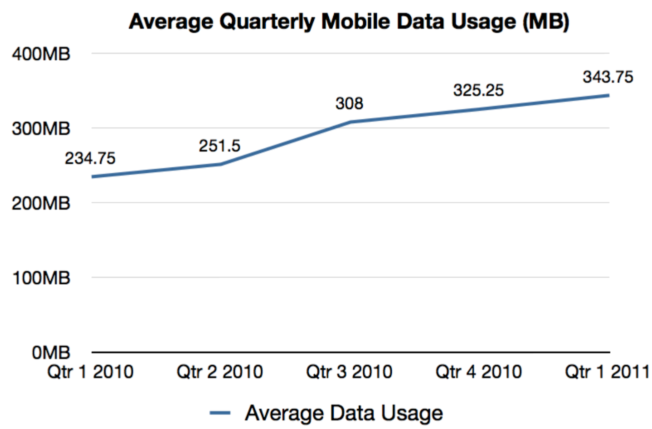

Figure 4: Average Quarterly Mobile Data Usage compiled from data in Nielson (2011). Sources averaged are Android OS, Apple iOS, Blackberry OS and Windows Mobile.

At the same time that data usage is increasing, the cost per megabyte of downloaded data has decreased. In the same Nielson report, the cost per megabyte of data downloaded dropped from US$0.14 at the beginning of 2010 to US$0.08 at the beginning of 2011. This demonstrates that as the cost of data is decreasing, consumption is increasing.

Domain Drivers (Data Transport)

In summary, the domain drivers for data transport energy consumption have been identified as:

- as the volume of data increases, the transport energy increases
- transport accounts for a large percentage of relative power consumption for transporting data to the cloud (Figure 3a)
- there is a relationship between cost of data, and volume of data downloaded

As more devices become available to access the data, the data can be downloaded multiple times. In the next section, we examine the effect of new devices on energy consumption.

3.4. Device Adoption

This category has been discounted in both the CDP (2011) analysis, and the Google (2011) response. A footnote from the Google analysis says:

> *There is no appreciable difference in client energy, since the user is usually not changing the device they use to access email.*

This statement appears somewhat at odds to their deployment strategy for accessing Google Mail through mobile devices, which shows 16 methods available to access Google Mail via a smartphone. This is included in Table 3 to demonstrate that there are multiple methods on multiple devices to access the same data.

Table 3: Supported phones for Google Mail on various platforms shows that there are a number of ways to access their cloud email service on a mobile phone. Data from Google (2012).

Access Choice	Android	Blackberry	iPhone	Nokia S60	Windows	Others
Web App	Yes	Yes	Yes	Yes	Yes	Yes
Native App	Pre-installed		Yes			
Sync via IMAP	Yes	Yes	Yes	Yes	Yes	
Google Sync			Yes	Yes	Yes	

We propose that the availability of cloud-connected devices has a strong adoption effect. Data that can be accessed through a number of devices encourages interaction with the cloud-based services and, in turn, encourages the creation of more cloud-based applications.

A user that accesses a cloud-based email service may simultaneously be accessing that data on a number of devices, such as their desktop, laptop, tablet and phone. If this were the case, data that travels between the cloud and the user could increase up to fourfold. The availability of new devices, and the lowering of data costs, are significant reasons for adoption of cloud-based services. For this reason, device adoption should be included in the dynamic model.

Domain Drivers (Device Adoption)

The domain drivers for device adoption energy consumption have been identified as:

- as the demand for cloud computing increases, an opportunity to improve data utility opens up. As the device gap is filled, it reinforces the demand for cloud computing
- as more devices are used, the energy required to power devices increases

In the next section, we combine these drivers from the four domains into a causal loop diagram.

4. Model Mapping

In this section, the relationships between the four domains has been mapped into a model boundary chart (Table 4) and causal loop diagram (Figure 5).

The model boundary chart in Table 4 states the boundaries that were considered when constructing the proposed model. We have approached our study to look at the energy scenario for a business that is not influenced by market and economy-wide fluctuations. This is a deliberate, as it matches the situation described in CDP (2011), which could also be described as a 'business-as-usual' scenario.

Table 4: Model boundary chart for energy consumption in cloud computing by domain

Endogenous	Exogenous	Excluded
On-site energy consumption Off-site energy consumption Data transport energy consumption Data consumption Device energy consumption Investment in technology Improvement in technology capacity Business decision pressures Demand for services Demand on infrastructure Cost of energy consumption Work practices Technology expectations	Population Competition Market share Employment cycles Green Accounting GDP Profit and loss Government Policies Energy Price Fluctuations	Material consumption Production energy Technology leaps Environmental constraints Company growth/decline Data access method variability

The definition of this model boundary helped in mapping the causal loop diagram, which is our dynamical hypothesis. This is shown in Figure 4.

There are two aspects of the causal loop diagram that have not been explained in Section 3. The *expected utility improvement* drives provides opportunities to improve workstation and data utility. This utility expectation is analogous to the expectations in Moore's Law (Moore 1965): that the performance of computing will continually improve.

The causal link between *Off-site energy consumption* and *Energy efficiency programs* describes a link that does not directly exist. It may be the case that this occurs as an increased financial cost, but we were unable to find any relationship between energy consumption and cost of the cloud service. This may become more obvious if a price on carbon is introduced, or if the direct accounting of energy was required to be reported. For this reason, it is indicated with a delay.

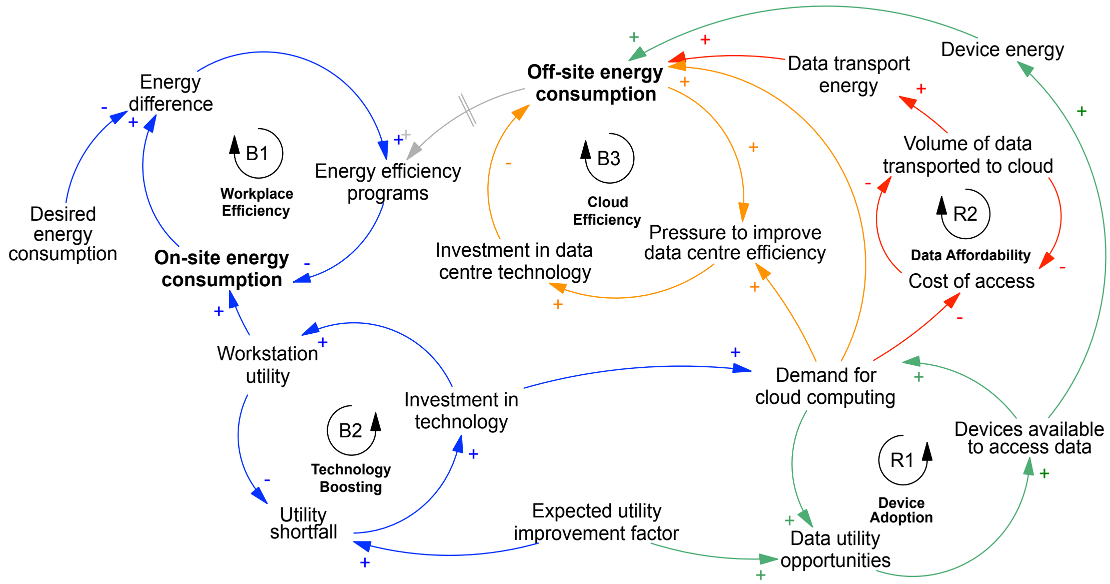

Figure 5: Causal loop diagram of domain, as mapped from trend data in Section 3. Expected utility improvement factor and the relationship between Off-site and On-site energy consumption is discussed below.

Figure 5 maps our dynamical hypothesis in a causal loop diagram. We discuss the relationships in the following section.

5. Model discussion

Our dynamical hypothesis leads us to observe that using the cloud appears to only add to the total energy consumed in the system. In this section we look at the implications of the causal loop diagram, and suggest possible 'what if' scenarios. The relationships have been mapped to reflect the behaviour described in the trends in Section 3.

The *workplace efficiency* loop (B1) reduces the on-site energy consumption independently of the technology that the business has in use. The *technology boosting* loop (B2) seeks to meet the desired level of utility, which is continually increasing as the *expected utility improvement factor* pushes expectations higher. As the workstation utility increases, the on-site energy consumption goes up.

The *cloud efficiency* loop in data centre efficiency (B3) reflects the attempts to improve data centre PUE. Demand for cloud computing increases the pressure for *cloud efficiency* as the demand grows.

The reinforcing behaviour mapped in the *device adoption* loop (R1) drives energy up through an increased popularity of new devices. As the *data affordability* loop (R2) drives the cost of access down, the energy required to power increasing data transport drives the energy consumption up.

We expect that the energy consumption will increase significantly in the whole system in this dynamical relationship. In our dynamical hypothesis, the *on-site energy consumption* has no feedback from the other loops. There may be a relationship that reduces on-site energy consumption through the removal of installed servers; however, we have not been able to find sufficient evidence that this is significant.

We were also unable to find any evidence to suggest that an increase in off-site energy consumption has any effect on the decisions made in on-site computing. This could be the scenario if the use of the cloud lengthened the time that they replaced workstations, or if they were required to account for off-site energy.

There are a number of 'what if' scenarios that could lead to a comparative reduction:

- if the adoption of devices, which are generally of lower power consumption than workstations, led to the discontinuance of the trend for desktops with large screens.
- if the processing power shifted to the cloud and the workplace policy on machine replacement took place over a longer time
- if the off-site energy footprint fed back to the decision-making process in house. This could be through a requirement to publish or report the off-site energy consumption.

When a numerical model is constructed, these 'what if' scenarios should be considered.

6. Conclusions

We investigated the trends around workplace computing and associated shift to cloud computing. We found that previous studies did not consider all of the domains required to properly account for the total energy consumption. Our dynamical hypothesis embodied in our causal loop diagram indicates that there are reinforcing loops that encourage increased data consumption, and an increase in the number of devices that can access the data. Further work is required to understand this dynamical relationship, but there is a likelihood that the total energy consumed by moving applications to the cloud actually increases, in contrast to reports that suggest otherwise. Further work to investigate this claim will be done through a numerical model, which is the next stage of this research.

7. Bibliography

Apple Inc., 2012. *Tech Specs.* Accessed 22 February 2012. <http://support.apple.com/specs/#desktopcomputers>

Baliga, J., Ayre, R.W.A., Hinton, K., and Tucker, R.S., 'Green Cloud Computing: Balancing Energy in Processing, Storage, and Transport', *Proceedings of the IEEE,* Vol. 99, No. 1, January 2011, pp149-167.

Bray, M., 2006, *Review of Computer Energy Consumption and Potential Savings,* White Paper. Dragon Systems Software Limited.

Buyya, R., Yeo, C.S., and Venugopal, S., 2008, 'Market-Orientated Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities', *10th IEEE International Conference on High Performance Computing and Communications, 2008.*

Carbon Disclosure Project Study 2011, *Cloud Computing - The IT Solution for the 21st Century,* Verdantix, 2011.

DEHWA, 2009, 'ENERGY STAR Computers in Australia 2009', Department of Environment, Heritage, Water and the Arts, Canberra.

EPA , 2007, *Report to Congress on Server and Data Center Energy Efficiency, Public Law 109-431*. Prepared for the U.S. Environmental Protection Agency, ENERGY STAR Program, by Lawrence Berkeley National Laboratory. LBNL-363E. August 2. <http://www.energystar.gov/datacenters>

IEA, 2009, 'Gadgets and Gigawatts - Policies for Energy Efficient Electronics', International Energy Agency, Paris.

Katz, R.H., 2009, 'Tech Titans Building Boom', *Spectrum,* Vol. 26 Issue 2. pp40-54

Koomey, J.G., 2008. 'Worldwide electricity used in data centers.' *Environmental Research Letters*. vol. 3, no. 034008. September 23. <http://stacks.iop.org/1748-9326/3/034008>

Koomey, J.G., Berard, S., Sanchez, M., Wong, H., 2009, 'Assessing Trends in the Electrical Efficiency of Computation Over Time', Final Report to Microsoft Corporation and Intel Corporation.

Koomey, J., 2011, Growth in data center electricity use 2005 to 2010. Oakland, CA: Analytics Press. July.

Google Inc., 2011, 'Google's Green Computing: Efficiency at Scale'. Accessed February 17 2012 <http://static.googleusercontent.com/external_content/untrusted_dlcp/www.google.com/en/us/green/pdfs/google-green-computing.pdf>

Google Inc., 2012, 'Gmail for Mobile'. Accessed 12 March 2012. <http://www.google.com.au/mobile/mail/>

Moore, G.E., 1965, 'Cramming More Components onto Integrated Circuits,' *Electronics*, vol. 38, no. 8, 1965, pp. 114–117.