
Early version, also known as pre-print

Link to publication record in Explore Bristol Research
PDF-document


University of Bristol - Explore Bristol Research
General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/pure/about/ebr-terms
Modeling and Assessing Variability in Use Phase Energy of Online Multimedia Services

Daniel Schien, Paul Shabajee, Dr. Mike Yearworth, Dr. Chris Preist

Address correspondence to: Daniel Schien, University of Bristol, MVB, Woodland Rd, Bristol, BS8 1UB, UK. Email: Daniel.schien@bristol.ac.uk

Keywords: Internet, life cycle assessment (LCA), information and communication technology (ICT), publishing, Monte Carlo simulation

Summary

Recent studies of the sources and extent of variability in life cycle assessments of the climate impact of information and communication technology (ICT) have highlighted a need to further investigate the uncertainty related to the energy consumption during ICT operation. We use an improved, more accurate model to analyze the energy footprint of the audience of a major online newspaper, which allows for the development of tools to monitor the energy consumption of ICT in real time. We identify the footprint of consuming news with various types of access devices and quantify the impact of behavioral parameters on the overall footprint. Previous estimations of national and global consumption assume average values for all subcomponents. Our results indicate that previous analyses based on average figures for laptops or desktop PCs predict values that are too high for national or global digital media power consumption. Additionally, we identify which components contribute most to the total use phase energy consumption and which should be focused upon in any life cycle assessment of digital services. We find the origin data center to contribute between 4 to 46% of the total energy budget when reading news articles and between 2 and 10% when watching video content. Similarly, we find that user devices contribute between 6 and 90% and 0.7 and 77% when consuming articles or video content, respectively.

DISCLAIMER: This is a pre-press version of the text. The definitive version is available at www3.interscience.wiley.com.

Introduction

The climate change impact of ICT has been studied by the academic community for some time, for example most recently by Malmodin et al. (2010) and Weber et al. (2010a); it is also attracting increasing interest from the public (Greenpeace 2012). Attributional life cycle assessment (LCA), the quantifying of environmental impacts resulting from the creation, use and disposal of a product or service, has played a key role in this analysis. A recent study by Weber (2012) of the sources and extent of variability in life cycle assessments of a server computer has identified the use phase energy consumption to be the most uncertain. The observation that the variability in the use phase is very high has also been made for end user devices (Beauvisage 2009).

The International Reference Life Cycle Data System (ILCD) Handbook distinguishes between variance as the degree of stochastic uncertainty in a single process within an LCA and variability as the single representation of multiple processes and systems with differing impact (EU JRC - Environment and Sustainability 2011). When faced with variability in a process flow, an LCA practitioner has a choice between a more detailed process model, which will require commensurately more acquisition of associated data, or a less detailed model, which takes an ‘average’ or ‘prototypical’ process and data set, concealing the underlying variability. The latter approach has the advantage of being easier, but may reduce the accuracy of the assessment and hide potential interventions.

In this paper we present an analysis of energy use in the delivery and consumption of online digital news content per individual, with particular reference to the variability that occurs in the delivery of the service. To do this, we have developed a model of digital service delivery that is significantly more detailed than prior art, and gathered associated data from new primary and secondary sources. The model
covers the use phase including the dynamic creation and delivery of content from a distributed set of data centers, transmission of content through routers, switches, and cable ('the Internet'), delivery of content through the user access network and consumption of content on user devices. Regarding the access network, we consider connectivity via digital subscriber line (DSL) modems in combination with domestic wireless network (WiFi) and third generation mobile networks (3G). The model we have developed, and much of the associated data, is of general applicability to digital services. Our analysis is focused on a specific case study: the provision of the multimedia website by Guardian News and Media Ltd (GNM). The functional unit we adopt in this paper is 10 minutes of content browsing, and we explore the impact of variability in content type (video and webpage), end user device (desktop, laptop, tablet, and smartphone), access network, geographical location and browsing behavior on overall energy use. Regarding the latter, we consider the impact of varying the speed of changing between web pages. We present results for a number of scenarios exploring the impact of this variability and variance within each scenario modeled, using a Monte Carlo approach, and identify how the significance of different components of delivery and consumption alters between scenarios.

This work contributes to state-of-the-art energy and carbon footprinting of ICT in a number of ways. Firstly, the model of energy use by digital services we present is both more detailed than previous studies and more complete in its ability to capture and distinguish between different usage scenarios. Secondly, using the primary and secondary data we have gathered, we present the first detailed analysis of a diverse set of scenarios for digital media consumption, and so update and complement existing studies of the environmental footprint of digital media. Thirdly, we present novel methodological advances in the modeling of digital services – specifically the use of Monte Carlo simulation to draw from alternative sub-scenarios rather than error estimates, and our discussion of the appropriateness of different allocation methods for different components of digital delivery. Fourthly, we present initial results that can be used for simplified analyses of digital services.

Our analysis of a large set of typical scenarios of digital online media consumption provides results that can be compared to several previous studies. The existing studies vary in the degree to which they apply bottom-up or top-down models for the majority of the life cycle processes. Our own model and others (Moberg et al. 2010; Chandaria et al. 2011; Williams 2011; Baliga et al. 2009) apply bottom-up models in as far as they calculate the energy consumption from the additive impacts for the most impactful processes per functional unit. On the other hand, in studies that mostly apply top-down models, such as Taylor and Koomey (2008) and Koomey et al. (2009), the measured or estimated total aggregate impact of an entire (sub-)system is related to the total number of functional units provided by the system. The result of both top-down and bottom-up allocation is an average value for impact per functional unit, yet only the bottom-up model contains data about the elementary life cycle processes. Our work draws from these studies but goes beyond them in several ways. Firstly, while they use aggregate data and assumptions regarding the average or typical user, we model the variability explicitly in a number of scenarios. Secondly, we use a Monte Carlo approach to account for both variance and variability within each scenario. Thirdly, we present a principled approach to the allocation problem as applied to digital services and make use of it in our model.

To our knowledge, ours is also the first study that relates the footprint of 3G networks to a functional unit of a media service. Although Scharnhorst et al. (2006) and Stutz et al. (2006) both provide an LCA assessment of a 3G cellular wireless network, their functional unit is that of a year’s mobile service. At this level of aggregation their results cannot be related to a single media service similar to our functional unit. Toffel and Horvath (2004), on the other hand, analyze the energy footprint of downloading newspaper content to a handheld reader via a mobile network but do so for a 2G wireless network. They reference the total energy consumption of the wireless network based on an LCA study from 1999 and relate it in top-down fashion to the total number of subscribers in the network. The most widely
A referenced top-down model of the energy footprint of Internet data transfer is presented by Taylor and Koomey (2008) in a study of the impact of web advertisements and we compare its results to those derived from our model in the discussion. The findings by Taylor and Koomey (2008) are applied by Teehan et al. (2010) in a top-down model that is used to analyze the total energy consumption in the U.S. for a variety of tasks, assuming that user behavior in the U.S. is similar to survey data from France 2005-2006. They do not capture the wide variability in the energy footprint resulting from individual user behavior and variability in the power consumption of devices.

This paper is structured as follows: in the next section we present the system model in abstract and identify suitable allocation approaches for individual subsystems. This section is followed by a description of the most significant model parameters. In the subsequent section we present the results of the analysis. In particular, we demonstrate variance and variability between scenarios of device, service type and access network combinations. We close with a discussion and conclusion in the final section.

**Models**

Broadly four categories of devices are involved during the use phase of a digital service. Firstly, data centers consisting mainly of servers but also of networking infrastructure provide the service content. In the case of GNM, this content is split between origin servers belonging to the organization itself and a number of third party servers. These either provide additional content such as advertising, or belong to content delivery networks (CDNs), which cache content in different regions around the world to improve service performance. Secondly, the devices that make up the edge and core networks of the Internet, used to transport data from its sources to the end user. Thirdly, the access network used to link the user’s device with the Internet, and finally the user’s device itself. Figure 1 captures this in a process model. Not included in our assessment is the impact of software development activities and editorial work. Our collaboration with GNM allows us to use primary data for many but not all processes.

Our functional unit is 10 minutes of browsing, during which we assume the user issues one or more requests for content. Each such request involves opening an individual uniform resource locator (URL) with the web browser. The energy consumption for each individual request is the sum of the consumption by the four subsystems in the delivery model. The energy footprint for the functional unit is the sum of the energy consumption of all requests issued during the time of the functional unit. Not included in this energy footprint is the energy consumption of other life cycle phases, notably the manufacturing of the devices. In the remainder of this section we will present the model in detail, starting with the allocation technique for shared IT infrastructure: the origin servers of the content provider, third party servers and the Internet. We will then look at the energy consumption of the network connection between the servers and the end user. Finally, we will
discuss the energy consumption of the end user equipment itself.

Allocation Approaches for Digital Products

A key methodological decision in LCAs is made during the allocation of environmental burden, which is defined as the act of “partitioning the input or output flows of a process or a product system between the product system under study and one or more other product systems” (ISO 2006). In this section, we consider alternative allocation approaches possible for digital products, and discuss their appropriateness in different situations. In the case of digital products, allocation is necessary for two reasons. Firstly, equipment use may be shared between multiple users, such as a content server providing web pages to many people, or a digital subscriber line multiplexer providing broadband connections to a number of households. Secondly, equipment may be used for multiple services, such as a physical server running multiple virtual machines, or a domestic laptop providing web access as well as email, playing music and many other applications.

The ILCD Handbook distinguishes between two approaches to allocation. The preferred approach, physical causality, allocates burden based on the share of some physical (or other) flow that is directly related to the environmental burden generated (EU JRC - Environment and Sustainability 2011). The second approach is to use some other relationship, such as economic activity.

We consider three approaches to allocation for digital services:

1. Data flow. In the case of digital services, there is no clear ‘physical’ flow to study, but there is a flow of data. Allocation can take place based on the share of data passing through an energy-using device. This approach is adopted by Lee et al. (2011) and Baliga et al. (2009), which considered both energy consumption by Internet routers and home access network devices.

2. Number of users. If a device is shared between services offered to a number of users, the energy can be allocated equally to each.

3. Service ‘attention’ time. If a user is using a given device for a number of services, energy usage by the device can be allocated based on the amount of time the user spends using the different services.

None of these cases correspond to the preferred approach of physical causality because energy usage of devices does not vary directly proportionally to data flow, users or services being used. We discuss this in more detail for each device in subsequent sections. To determine which approach to use we adopt a principle of allocating based on which of these is the limiting factor to device usage – namely the factor which, if increased, would first limit or degrade the quality of service. In the case of most network devices, this is usually bandwidth. Similarly, video, audio or image content servers such as those used by CDNs conduct relatively little computation and are limited by their capacity to transmit data at speed. In the case of a DSL multiplexer, on the other hand, which is used to provide access to the Internet for a number of premises, the limiting factor is the number of connections it can provide, and therefore the number of users it supports. Also, in the case of a web server, the limiting factor is the computational power required to construct pages rather than the speed of data it outputs. Finally, in the case of a user device, the limiting factor is usually – but not always – the user’s attention: the device could easily run more applications, but the user would only be able to make use of a limited number at a time. In addition to the limiting factor, we allocate along that dimension which, if changed – given current levels of typical utilization – would result in the most significant change in the energy consumption. For user devices, for example, one such choice is between data volume received and time of service consumption. In the case of most online multimedia services, and in particular online news, a reduction or increase in the device operation time will result in a much greater change in the energy consumption of the service than a change in the data volume transferred. In the remainder of this section we formalize the allocation of energy consumption for each system component beginning with the network devices.

Within the time interval of the functional unit, the user issues one or more requests for a URL corresponding to a separate web page. For each request a browser opens several connections to servers to retrieve all resources referenced in the
By application of the allocation principle of the limiting factor, explained above, we allocate to each network hop an amount of energy $E_h$ as a fraction of the total energy consumption of the device $E_T$ relative to the ratio of the data volume transferred for the individual connection $V_c$ to the total amount of data that the device serves $V_T$:

$$E_h = E_T V_c / V_T$$  \hspace{1cm} (1)

The total data volume that is served by a device is equal to its throughput of data $\varphi$ over time $t$:

$$V_T = \int \varphi(t) dt$$  \hspace{1cm} (2)

The throughput $\varphi$ depends on the utilization $u$ of the devices’ maximum throughput capacity $C$ integrated over time $t$. The utilization typically changes in a cyclical pattern over the time of day $T$ so that $\varphi(t, T) = Cu(T)t$. We assume that the utilization is relatively constant during short timespans of processing an individual customer request which is usually less than a few milliseconds. The energy per hop $E_h$ is allocated is then:

$$E_h = V_c \frac{P(u(T))}{C \cdot u(T)}$$  \hspace{1cm} (3)

During its operation a device draws a minimum of power $P_B$ when idle, which increases depending on the level of utilization $u$ of its components up to a maximum $P_M$. Formally, we define this with a function $\alpha(u)$ which ranges from 1 to $P_M / P_B$:

$$P(u) = \alpha(u) P_B$$  \hspace{1cm} (4)

Let $P_h$, $C_h$, and $u_h$ be the power draw per hop, its throughput capacity and its utilization, respectively. Additionally, all industrial grade power consuming network equipment requires cooling and power transformation, the losses of which are commonly denoted as power utilization efficiency (PUE). Then, the energy consumption for the data transport of the functional unit $E_R$ over the all hops in the route $r$ is defined as:

$$E_R = 2V_c \sum_{h \in E_r} \frac{P_{h_i}}{C_h u_{h_i}} P_{UE_{network}}$$  \hspace{1cm} (5)

where $P_{UE_{network}}$ denotes the overhead for cooling and power transformation and the factor 2 accounts for redundancy.

Regarding the servers, we distinguish between the origin servers (at GNM) and third party servers such as CDNs, advertisement and analytics services. At the origin data center at the GNM, the energy consumption of each origin server $E_S$ and the supporting networking infrastructure and storage devices in the data center $E_{net}$ and $E_{storage}$ can be measured directly. In the case of the origin servers at GNM, the energy per request was found to be relatively independent of its data volume. In other words, the limiting factor is the number of pages requested and we therefore allocate power consumption uniformly between all requests $n$ served during a measurement interval. Including the overhead for cooling and power transformation, $P_{UE_{origin}}$, which is an independent value for each data center, the energy consumption per connection to an origin server $E_o$ is thus:

$$E_o = \left( \sum E_S \right) + E_{net} + E_{storage} \frac{1}{n P_{UE_{origin}}}$$  \hspace{1cm} (6)

The power consumption of third party servers cannot be measured directly and must be estimated based on available public data. Below, we estimate a volume-dependent coefficient $\lambda$ of energy per data volume which is the only feasible allocation principle given the few available data on energy consumption of CDN servers. Each browser request for a URL results in multiple connections to third party servers. The total energy per request from third party servers is the product of $\lambda$ and sum of the data volume $V_R$ transceived for all connections of a request between a customer and the third party sources:

$$E_{3p} = \lambda V_R = \lambda \sum_{i \in \text{Connections}} V_{R_i}$$  \hspace{1cm} (7)

Next, we present the model for the mobile and wired access network. A mobile access network is composed of a set of base stations that provide a radio signal to smartphone client devices. The base stations are connected to the edge network via a
mobile service center. A base station provides signal coverage in a cell of a limited area. Multiple users can be connected to the same base station concurrently. In modern base stations, unlike network routers, power usage is affected by the load at a given time in two ways. At phases of sufficiently low utilization, power saving features can yield a substantial reduction of the base station’s power draw (Ericsson 2007). If the utilization exceeds that threshold, the energy elasticity of a base station becomes very low (Eunsung Oh et al. 2011), to the degree that in the context of this work we consider it static.

Mobile voice services require a constant bandwidth per user and thus the limiting factor is the number of users that can be served with the available base station capacity. Data services, on the other hand, vary significantly with regards of the bandwidth they consume and a single user can potentially consume the total available bandwidth of a base station (Julius Robson 2011), hence the limiting factor of a base station is its data throughput capacity. Given \( P_{BS}, C_{BS}, u_{BS} \) and \( PUE_{BS} \), a base station’s power draw, data throughput capacity, utilization and overhead for cooling and transformation respectively, and the total data volume per request \( V_R \), then the energy consumption by the base station per request \( E_{BS} \) is calculated as:

\[
E_{BS} = \frac{P_{BS}(u)}{C_{BS} \cdot u_{BS}} V_R PUE_{BS} \tag{8}
\]

There are several alternative domestic wired broadband access network technologies available, such as DSL, cable or fiber-optic, of which DSL is the most popular in OECD countries (OECD 2011). A DSL connection requires two types of equipment:

- A terminal unit such as a DSL Access Multiplexer (DSLAM) with a power consumption \( P_{TU} \) that provides Internet connectivity to a number \( N_{TU} \) of users. The DSLAM requires active cooling and high voltage power transformation the overhead of which we capture with a PUE factor \( PUE_{Net} \).

- A DSL modem at the user’s premises, which is frequently connected to a wireless router. We denote the sum of the power consumption of all network equipment at a given user’s premises as \( P_{CPE} \) and the number of users connected to them as \( N_{CPE} \).

DSLAM have a fixed capacity similar to Internet routers. The former, however, are hard-wired to a limited and fixed number of subscribers in contrast to Internet routers, which transport data from a changing number of users. We allocate the power consumption of the DSLAM uniformly among the subscribers and calculate the energy consumption based on service consumption time. We formally define the energy consumption of the access network \( E_{AN} \) as the product of the time of service consumption \( t_S \) and the sum of the power draw of each device type listed above:

\[
E_{AN} = t_S \left( \frac{P_{CPE}}{N_{CPE}} + \frac{P_{TU}}{N_{TU}} PUE_{Net} \right) \tag{9}
\]

The access network devices are connected to end user computing devices such as smartphones, tablets, laptops or desktop computers. We do not consider the power consumption of printers, scanners and other peripheral components that can potentially be powered on during browsing.

The energy consumption by the end user device is not primarily constrained by the data volume and the time for the reception of that data. With broadband, Internet text webpages and video buffers are often received within a few seconds, after which the connection to the server is paused. Instead, the energy consumption by the user’s device is primarily dependent on the time that users spend on consuming the content, hence we allocate energy by time. Note that user devices as well as home network equipment are often kept in idle mode during significant periods of time and continue to consume energy, the impact of which we exclude from our model but consider in the discussion.

The base power consumption of end user computing devices \( P_{BUD} \) fluctuates with the level of usage as described above in eq. 4. Carroll and Heiser (2010) find that this also is the case for smartphones. Hence, the energy consumption of the user device \( E_{UD} \) is the product of its power consumption and the time during which the service is consumed \( t_S \):

\[
E_{UD} = \alpha P_{BUD} t_S \tag{10}
\]

**Model Parameters**

In the following section we present a parameterization that allows calculating an energy footprint for a media service provided by Guardian News and Media. The main model parameters are
power consumption values, throughput capacity and operational use time of servers, network devices and user devices. We will look at each parameter in turn and discuss possible sources for values and variability.

Firstly we consider the GNM data center. GNM operates virtualized blade servers that are arranged in a tiered array of 25 blades, each of which has a power consumption varying between 140 and 300 watts (Beckett and Bradfield 2011) from base to peak load. Apportioned to the relative number of monthly visitors, this number of servers is similar to that reported for the German online magazine stern.de (heise Verlag 2011). The average number of pages served per second from all servers ranges from about 40 to 200 with a trough during the early morning (Wood 2012). Also from internal data we know the load of the servers typically varies between 15 and 30 percent, which is in agreement with typical utilization of data centers as presented by Barroso and Hölzle (2007). Network and storage equipment in the data centers are often shared between independent parts of organizations which necessitates allocation decisions. In order to assess the impact from varying the allocation of this equipment we sample this contribution from a triangular distribution between a minimum of 10%, a maximum of 30% and a mode value of 15% overhead each. For the additional overhead from cooling and power transformation $PUE_{origin}$, we apply a distribution based on the values from Bertoldi (2010) between 1.25 and 2.86 (see appendix figure 12 for more details).

The GNM commissions several CDNs, of which Akamai delivers the largest data volume for pages without video content. Akamai reports an annual carbon footprint of 8 kg CO2-eq/Mbps (Akamai 2010) and a carbon intensity of 0.59 kg CO2-eq/kWh, an average of 40% idle power consumption (Energy Star 2011), and an average utilization of servers of 80% (Akamai 2010). Based on these values, we estimate the energy consumption per data volume to be $2.14 \times 10^{-6}$ joules/bit ($J/b$). This is roughly five times less than the value reported by Google for Youtube servers ($1.33 \times 10^{-5}$ $J/b$, assuming bit rate of 900 kilo bit per second (kbps) (Google 2011) and more than five times the value of $4 \times 10^{-7} J/b$ assumed by Chandaria et al. (2011). Given the discrepancy between those values we do not make additional assumptions about the possible variation of energy efficiency from changing utilization over the time of day.

A note regarding the usage of units for energy consumption: in the existing literature, values of energy consumption have been presented in joules and watt hours. Given that the numeric values per bit are already very small, we use joules in this text.

Our network model, defined in eq. 5, is a bottom-up model that needs to be parameterized with the number of hops in the connection between a server and the user. The alternative modeling approach of top-down modeling, in the case of the core network, estimates the energy efficiency of a single service from the total energy consumption of the entire network and apportions it to all services delivered. Both approaches arrive at different values, and we discuss the discrepancy between top-down and bottom-up models in more detail below. During the Monte Carlo simulation, we evaluate the impact of these different assumptions by sampling from a triangular distribution. The minimum and mode values of 4.5 and 10.5 joules/megabit ($J/Mb$) are based on a bottom-up model which we present in detail elsewhere (Schien et al. 2012). For the maximum value we apply a value of 36$J/Mb$ based on a top-down model (Malmolin et al. 2012).

The power consumption per subscriber by the DSLAM was assumed to be around 2W by several studies (Aleksic and Lovric 2010; Lee et al. 2011; Baliga et al. 2009). In addition to that prior research, our own measurements also find that the power consumption of DSL modems is approximately 5 watts. Separate WiFi routers have a similar power consumption, as measurements for the new Energy Star rating of small network equipment indicate (Energy Star 2012a). We assume a home setup of a single wireless router and single DSL modem with both being actively used for the same time as the end user devices.

The energy efficiency of cellular wireless networks varies strongly with the allocation of the energy consumption for cell subscription required to receive calls. In our Monte Carlo simulation we apply a triangular distribution with a maximum value of 328 $J/Mb$, which is based on uniform allocation of total power consumption to data packets. For the mode and minimum of the distribution we apply values that are based on the allocation of the instantaneous
power consumption of the base station and the data rate per subscriber. We apply 54.36 J/Mb as the mode and 13.29 J/Mb as the minimum, which are based on a power consumption of 460W per subscriber and a data rate of 11Mb/s and 45Mb/s for High Speed Packet Access, a third generation cellular network evolution from Deruyck et al. (2010). The minimum and mode value also include an overhead of 1.3 to account for the energy consumption of the remaining parts of the cellular network in addition to the mobile base station. This is based on the yearly energy consumption of 4.177GWh for the whole network of the German mobile network operator Vodafone for operation of 224.000 base stations (Vodafone 2011), resulting in an allocated average power draw of 2129W per base station which is about 30% more than the average nominal power consumption of base stations operated by Vodafone Germany (Zwemke 2012). We assume that the energy efficiency of mobile networks is similar between OECD countries although no systematic study exists.

We distinguish between the following classes of end user devices: smartphones, tablets, laptops, and desktop computers. For the laptop and desktop computers (including monitors), distributions of power consumption are based on data from the Energy Star measurements (Energy Star 2011, 2012b). These extensive lists contain power measurements of several thousand energy efficient devices that were awarded the Energy Star rating. They do not represent the relative popularity of these devices.

On top of the power consumption in active idle mode (non-standby), the computational complexity of programs introduces a dynamic portion of power consumption. The relative and absolute magnitude of this dynamic power consumption depends on several parameters: for example, the specific device and – in the case of browsing online news – the amount of JavaScript embedded in a page or the video codec used. Yet, systematic research of the influence of these parameters does not exist. We conducted a scoping experiment on a single, modern Energy Star-rated laptop and found no statistically significant variation from idle power when browsing text, hence, we set $\alpha = 1$ in eq. 10. In the consumption of video, however, the same experimental setup found a significant increase in power consumption. In our model, we assume the power consumption of devices increases by 15%, hence, we set $\alpha = 1.15$, which is similar to values reported in Somavat et al. (2010).

Based on GNM data, we apply an empirical distribution of the duration that users spend reading or watching the news. The distributions can be found in the appendix in figures 9 and 10. The distribution of news reading is heavy tailed and has its average at one and a half minutes. Video content is being watched for approximately two minutes on average.

**Method**

We have conducted simulations of a number of different scenarios of users accessing GNM digital services to explore the impact of variability on the energy footprint. Each scenario has a specific user device, access network technology and service type associated with it. The service type can be either an HTML web page including text, images and gif animations, or HTML with embedded video content. The functional unit for either service is 10 minutes of browsing. The average duration spent per website is 90s for reading text and 121s for videos. The precise value of the duration per page is randomly sampled during each iteration of the Monte Carlo simulation. We simulate the most popular user device technologies: smartphones, tablets, laptops, and desktop PCs. We exclude some exotic combinations of local access network types and user devices such as the combination of wired connection of a phone or a tablet to a DSL modem, but we include the simulation of smartphones connecting to a wireless home network. We only simulate mobile network access for phones, tablets, and laptops. We do not consider mobile access with a laptop in combination with an external screen.
For each scenario, we conduct a Monte Carlo simulation of 100,000 runs. This figure was determined by experiment to ensure convergence of average total energy consumption to within 0.1% overall for the same scenario. Each run draws from distributions based on both variance and variability within a given scenario. Variance is handled in the usual way, by using distributions around a mean value based on data quality factors and correlation between different secondary data sources. We give details of the distributions used in the appendix. Our approach to handling variability is novel; instead of a statistical distribution around a mean, we make random draws from a representative population of discrete observed values. For example, in the case of time taken to view a given web page or video, we make draws from a distribution generated from actual usage data provided by GNM. Similarly, the geographical location of the end user and the time of day accessed draw from distributions of GNM primary data. In the case of end-user device power consumption we make draws from a population of potential device models, each with an associated power consumption. Again, details are provided in the appendix except where commercially confidential. In all cases of variability represented in this way, our model allows values to be fixed to give results for a specific sub-scenario – for example, modeling a user accessing the service from Boston with an iPhone at 6:00 P.M. GMT.

Results

We now present results of our simulations for the scenarios we explored. We start with the presentation of average absolute values of energy consumption, broken down according to contributions by different system components. We then show which components affect the total allocated individual power consumption most and explore this influence in more depth. Figure 2 shows the average energy consumption for the different scenarios. The error bars indicate the 25% and 75% percentile of the sample distributions. Both figures share the same vertical scale. The energy consumption varies widely between the scenarios highlighting the need to take the particular combination of device types, local access networks and service type into account. The average energy consumption for consuming video is higher than the energy for consuming text. Not surprisingly, in the case of reading articles, the scenarios with a desktop computer arrive at the highest total energy consumption with an average of 96J. The least energy is consumed when reading articles during 10 minutes with a smartphone over a cellular wireless connection with 9J energy consumption. In the case of consuming video content, the scenarios with a cellular wireless connection rank highest where the 3G network alone contributes 121J, which is up to 83% of the total energy consumption. The energy consumption by core and edge networks is smaller but – at 15J – not insignificant. The energy consumption of scenarios

![Figure 2 Average energy consumption for 10 minutes of news consumption by system components for selected combinations of access network and user device type.](image-url)
with handheld devices (smartphones and tablets) is dominated by the network and servers while the energy consumption of scenarios with laptops and desktops is primarily dependent on the user device power consumption. The total energy consumption varies substantially between scenarios. The full numeric values are presented in the appendix in table 1 and 2.

Figure 3 shows histograms of selected sample distributions within the 2.5th and 97.5th percentile. The histogram for tablet scenarios is almost identical from that of the smartphone and is not shown. Considering the non-video scenarios, those with a smartphone show a much smaller degree of variability compared to those with PCs. This is mainly due to the larger variability in the power consumption of PCs in comparison to smartphones or tablets. The histograms also show that there is a clear distinction in energy use between laptop and desktop computers, with laptops only using more energy than desktops in 2.13% and 2.42% of the scenario samples for text and video respectively.

Following Weber (2012), we use a Spearman rank analysis over several scenarios, varying the access network type and the service type, to determine how different parameters of the model affect the final result. It generates coefficient values $\rho$ between 1 and -1, with $\pm 1$ indicating perfect correlation or anti-correlation, and 0 indicating no correlation.

![Figure 3 - Histograms of total energy by all subsystems for selected combinations of access network and user device types and consumption of 10 minutes of news with and without video content.](image-url)
Figure 4 shows the average values of those correlation ranks between scenarios of consuming video or text depending on the local access network type with an absolute value greater than 0.1. The top row shows the average values over scenarios with WiFi access, namely smartphones, tablets, laptops, and desktops and, similarly, the bottom row shows the average correlation values of scenarios that include 3G mobile access, namely smartphones, tablets, and laptops. The whiskers represent the maximum and minimum values for between the scenarios within one analysis set. For example, in the scenario of consuming video over WiFi in figure 4, the average correlation between the power consumption of the user device and the total energy consumption is 0.42, yet in the specific case of a smartphone (minimum) it is 0.04 and in the scenario of a desktop (maximum) it is 0.92. Importantly for video services consumed on handheld devices, the total energy consumption depends most strongly on the access network rather than the user device. Also, for the video scenarios the access network is much more relevant to the total footprint than it is for the text scenarios. Not surprisingly, the correlation between server utilization, expressed by pages per second, and the total energy consumption is higher in the ranks for 3G mobile access than in those for WiFi, since the former do not include desktop scenarios. Negatively correlated coefficients indicate inverse correlation of components; for example, the lower the utilization of the origin servers the higher the total energy consumption. Also, duration appears negatively correlated with the total energy footprint as it is inversely proportional to the number of repeated page requests submitted within the 10mins. When connecting with 3G the shared access network has a stronger impact on the total power consumption than the home networking equipment when connecting with WiFi. In the case of watching video, the data volume is directly dependent on the duration of the service consumption. When consuming text only, the
data volume has much less impact on the total energy consumption.

Discussion

Analysis

In this section we discuss our results in the context of previous work modeling the use phase energy consumption of digital media. We compare the quantitative results with those of other authors, where there is overlap of the models, and explore the reasons for differences. The energy per bit varies between cellular wireless and wired access network connections and also depends on the data volume of the service consumed. The average values are 17 J/Mbit for the edge and core network and 132 J/Mbit for 3G text and 9J/Mbit for video over DSL and WiFi (compared to an average 577 J/Mbit for text as a consequence of the allocation model). Williams and Tang (2011) allocate power consumption only for the duration of the data transfer, resulting in an energy efficiency per bit for wired connections of circa 4 J/Mb, which is a fifth of our results. For the servers, on the other hand, they arrive at a much higher energy footprint per user for browsing web pages by assuming that a server is occupied during 50% of the duration the user spends reading a page, while we use primary data from GNM showing that web servers complete page requests in sub-second time intervals. Figure 6 compares results for energy per bit on the Internet and access network in two of our scenarios with the results from the earlier works. In this table we include only the energy consumption per bit by the Internet and access networks, which we assume to be independent from the type data transferred. Baliga et al. (2009) estimate a value slightly lower than our minimum assumption of 3.23 · J/ Mb for the sum of edge and core routers and optical transport. While they assume 100% utilization of Internet routers and we assume between 12 and 25% (TeleGeography 2005), our measurements of hop count per route (from 6.5 to 15) is on average lower than their assumed value of 14. The difference regarding the access network power consumption is the result of a different allocation model. They allocate by throughput capacity; we allocate by time. We argue in the model section above why we believe time to be a more appropriate metric. Moberg et al. (2010) do not take into account the energy consumption by servers. Idle energy consumption is then apportioned relative to the duration of service use.

Chandaria et al. (2011) do not take account of the energy consumption of the Internet in their calculations. Their result for the wired access network is 11J/Mb and ours is 9J/Mb. This similarity is accidental. They take into account the idle power consumption of the DSL modem based on the assumption that it is used for 10.75 hours per day and idle for the remaining time and allocate it to the active use time similarly to Moberg et al. (2010) We on the other hand include a wireless network router besides a modem (both 5 watts) but do not account for idle power consumption. The reason why we do not include idle power consumption of user equipment in this assessment is the current lack of systematic studies of this important factor to the energy consumption at the user premises. This problem is further compounded by allocation questions of the idle power consumption. Even though every new generation of mobile networks brought a decrease of the energy consumption per bit of data, the total power consumption of base stations increases with their total throughput capacity (Manner et al. 2010). This, together with higher bandwidth usage by mobile services (Cisco Newsroom 2012), means their relevance will grow. Our assumptions regarding the energy efficiency of mobile data transfer overlap with those by Toffel and Horvath (2004). They relate the total energy consumption of a 2G mobile network to the total number of subscribers in the network and determine
a power draw per minute of 840W. This average power draw is then applied to the transmission of data which is assumed to endure 60 seconds over a 56kbps modem. The resulting energy efficiency of \(1.5 \times 10^{-2}/b\) is circa two orders of magnitude higher than our average values. This discrepancy mainly results from outdated values for the utilization of mobile networks and from using the energy footprint of voice service to calculate the footprint of a data service.

In their top-down study, Taylor and Koomey (2008) find the energy footprint per data volume to range between 9 and 16kWh/GB. This figure has been referenced and updated by several other studies extrapolating using a trend identified by Taylor and Koomey. Weber et al. (2010a) use this in a comparison of the environmental impact of different methods for delivering music and assume a value of 5-7kWh/GB. Preist and Shabajee (2010) estimate an upper bound on future global energy use for the provision of media services and extrapolate to a value of 4kWh/GB from Weber’s value. In order to compare this value to the results derived from our model it is useful to consider the results separately for servers and the network in the way that Moberg et al. (2010) performed their calculation. For the data transport they also apply Taylor and Koomey’s (2008) values, excluding the contribution of servers, to give a value of 3kWh/GB. Taylor and Koomey (2008) take the energy consumption values from a study by Roth, which accounts for network components used in a commercial context (Roth et al. 2002). Roth’s inventory is now severely outdated but in order to compare Taylor’s values with ours it is necessary to analyze this data in more detail. They distinguish between several device types, among which only the WAN switches and routers map to our model of the public Internet. They calculate the energy consumption on total shipments of network devices, which include ISPs, commercial intranet, and domestic deployments, and accordingly their results are likely to overestimate the energy consumption when applied to calculate the power consumption of the public Internet. Assuming that the three device categories hubs, routers, and WAN switches contain the devices which we consider the public Internet, then the energy consumption of the Internet would only account for 14% of Taylor and Koomey’s values.

Applied to the latest values extrapolated by Preist and Shabajee (2010), this would result in an energy footprint for the Internet of \(2.4 \times 10^{-4}/b\), which is roughly 14 times higher than our values. This discrepancy is either the result of an overestimation on the side of Roth et al. (2002), an underestimation of the network traffic in the Internet or a severe underestimation of the number of devices and their energy consumption in our bottom-up model.

Data Quality

Data on GNM server energy consumption and duration of service use was provided as primary data by GNM, so it is of high quality. For energy consumption by third party servers, we use a figure estimated from annually aggregated publically available emissions data from Akamai, one of the largest content delivery networks for the media industry, and use this for all third parties. Other CDNs are likely to have similar figures for data intensive streams, but this is likely to be an underestimate of servers of less data intensive content, such as advertising content providers and data analytics servers. Our model of the Internet distinguishes between edge and core routers. For each class of router, we have a number of data points from manufacturers’ specifications and peer-reviewed literature, which we use to generate a mean value and statistical distribution. Our model of the access network uses a similar approach. It omits certain equipment which is operated by some but not all ISPs – for example VPN connection servers between the access network provider and the Internet ISP or Remote Authentication Dial-In User Service servers – due to lack of publically available data. Although we believe that this is acceptable as a lower bound for the access network power consumption and that inclusion would increase the portion of the access network without significantly altering the result of the assessment, further research would benefit from transparency of ISPs in this area. Data on energy use of end-user devices comes from Energy Star and so can be considered primary data. The relative quality of the different data points was used in determining the range of variance of parameter distributions used within the Monte Carlo simulation described below.
Implications and Applications

The use of aggregate figures and assumptions about typical user behavior may be adequate for environmental accounting and reporting purposes, yet it can conceal insights into the impact of variability on an energy footprint that can be used for a number of other purposes. As proposed by Weber et al. (2010b), a more detailed model can be used to support real-time feedback to a user about the energy and climate impact of their behavior. We discuss how distributed systems technology can be used to support this in Schien et al. (2011). Such a model can also be used to support the environmental strategy of an organization wishing to reduce the footprint of its digital services. It can be used to assess different interventions for their potential impact, and support ‘design for environment’ of digital products. For example, in the case of the website analyzed here, a number of measures are suggested from the scenario results and the Spearman rank analysis. These show that choice of user device is the most significant factor in determining the use phase energy footprint of the service. This suggests that encouraging a move to smartphone and tablet access will have a significant positive impact. This can be done through the provision of apps that enable enhanced experiences on such devices, provided that such a move does not stimulate additional purchases or an increased upgrade rate of such devices. It shows that data transfer of video content has a significant energy use on the 3G mobile network, but less so elsewhere. Hence a strategy of reducing the resolution of video would be appropriate for mobile devices, but unnecessary for other devices. If the browsing time of users is assumed constant, the model also shows that the duration of time spent on a page is inversely correlated with energy consumption, particularly if that page is of text or images rather than video because the user is looking at multiple pages, the delivery of each of which adds to the energy consumption. This suggests that focusing on the design of web service and content to enable users to easily get to content that is most of interest to them, and ensuring it is of sufficiently high quality that they want to stay with it, is beneficial in terms of both the energy footprint and as a business strategy.

Recently, data center energy consumption has received heightened public attention, for example by (Cook and Horn 2011). Though increasing awareness of this issue is justified, our analysis, together with that of others, shows that for many organizations energy use by user equipment and the mobile network are bigger contributors to the service footprint. Data centers are assuming the role that plastic bags have for supermarkets, receiving attention disproportionate to their relative contribution of environmental burden compared to other parts of the retail business. It is important that the analysis of the impacts of IT, and the means to mitigate these, takes a view of the entire system.

Our work also highlights the importance of allocation techniques that are in accord with the technical functionality and usage of the system under study, and this is particularly challenging in the area of distributed IT systems. The choice of an allocation technique can have a significant impact on the results of the LCA. Our work makes a contribution to the debate of how best to do this, although we do not claim that we have provided the definitive answer. In particular, we allocate all energy of a user device to one function – namely browsing a website – while the user is carrying this out, even though the system could be carrying out other functions simultaneously. For example, it may be playing music. And it is likely providing instantaneous availability of services such as email, Internet telephony or instant messaging chat. The question of how best to allocate user device energy between these requires further work. Furthermore, a user device has periods when it is consuming energy on standby, or is on but not providing any active functionality. How best to allocate the energy used during these periods between the various functionalities it provides is also a question meriting further exploration.

Beyond the scope of this paper, it is relatively straightforward to extend our analysis to cover greenhouse gas emissions. The model identifies the different locations where electricity consumption takes place in the use phase of a service. This can be combined with national and regional carbon intensity figures, where they exist, to give a more precise estimate than would be possible using a single global or national intensity figure. Our work
can also be extended to allocate energy and carbon emissions associated with manufacturing the equipment to the digital services. This is obviously an important part of the overall footprint, and should be accounted for when making comparisons with alternative delivery methods of news content, such as paper-based.

The global IT system is responsible for the consumption of 3.9% of electricity (Malmodin et al. 2010). A significant amount of effort has been put into reducing energy use of individual components – such as laptops and data centers – motivated by eco-efficiency and cost savings. While this is valuable, it does not address the energy and environmental consequences of design decisions taken by the various parties involved in providing services across the internet. The complexity of the business ecosystem involved in such services means that a design decision by one can have energy (and therefore environmental and cost) implications on many others. Similarly, choices by the end user have effects throughout the system, and those choices are influenced by the service provider. The energy model presented in this paper is detailed enough to allow assessment of the implications of such decisions and choices. It allows the systemic approach that characterizes industrial ecology to be applied to the IT business ecosystem in a number of ways.

Firstly, such a model can be used to support real-time feedback to a user about the energy and climate impacts of their online behavior, as proposed by Weber et al. (2010). We discuss how distributed systems technology can be used to support this in Schien et al. (2011). While this may be of interest to some users, we do not see this as likely to lead to significant energy reduction without action by the service providers. Service providers can use our model to assess the effect of possible user trends on energy use by their service, and use this to consider which trends to encourage and which to discourage. For example, in the case of the website analyzed here, the Spearman rank analysis shows that choice of user device is the most significant factor in determining the use phase energy footprint of the service. This suggests that encouraging a move to smartphone and tablet access will have a significant positive impact. This can be done through the provision of apps that enable enhanced experiences on such devices, provided that such a move does not stimulate additional purchases or an increased upgrade rate of such devices.

Secondly, such a model can be used to assess the impact of decisions by designers of a digital service on the energy use of that service across the IT system, and propose design modifications that result in reduced energy use. For example, our analysis shows that data transfer of video content has a significant energy use on the 3G mobile network, but is less when transferred over other networks. Hence a possible design intervention would reduce the resolution of video automatically when the service provider detects the service is being delivered over 3G, but leave high resolution at other times. Such an intervention, if widely adopted among video service providers, could significantly reduce load on the 3G mobile network, and hence associated energy use, environmental impacts and costs. Beyond the GNM website analysis presented in this paper, our approach can be used to evaluate other IT design and architectural decisions from an energy perspective. For example, Apple’s iCloud music match service fingerprints songs of a user’s music collection locally and adds the identified songs to the cloud library from the existing cloud repository and thus avoids redundantly uploading terabytes of music files (Schien 2012). Another intervention that can be evaluated with the model is increasing outsourcing of data from the servers of a host such as GNM to the CDNs, who can serve content more efficiently and benefit from economies of scale at the same time as reducing bandwidth in the core network, realizing additional energy savings.

More broadly, such a detailed model can be applied to questions of ‘virtual industrial symbiosis’. Certain internet architectures used by service providers, such as the peer-to-peer architecture used by the Spotify music streaming service, use ‘waste’ compute cycles on customer machines to deliver content on other machines. The prime motivation of such architectures is cost reduction (by avoiding energy and infrastructure) at the service provider. Our model could be extended to allow assessment of such architectures to determine if they do reduce energy consumption across the system, or simply move the energy burden away from the service provider.
With the increasing pervasiveness of digital technology, the increasing sophistication of online services, the increasing energy consumption by IT and the increasing complexity of the business and technical systems which deliver them, it is necessary to go beyond local optimization of energy use and environmental impacts, and adopt a systemic perspective to mitigation as advocated by Industrial Ecology. By providing a model of digital services detailed enough to explore the impact of design interventions on energy use across the system, we facilitate the adoption of such a perspective.

Acknowledgements
This work was conducted as part of the SYMPACT project, funded by the RCUK Digital Economy and Energy programs (EPSRC EP/I000151/1). The authors would like to acknowledge the contribution of time and data made by Guardian News and Media, particularly Matthew Malthouse, Christopher Hodgson, Jo Confino and Stephen Wood. For her suggestions and for proof reading the article we also want to thank Elaine Massung.

References


Informatics for Environmental Protection. Shaker Verlag.


About the Authors

Daniel Schien is a Research Assistant at the University of Bristol’s Faculty of Engineering, School of Computer Science (UK).

Dr Chris Preist is a Reader in Sustainability and Computer Systems at the University of Bristol.

Paul Shabajee is a Research Fellow at the University of Bristol’s Faculty of Engineering, School of Computer Science (UK).

Mike Yearworth is a Reader in Engineering Systems in the Department of Civil Engineering at the University of Bristol (UK).