EnviroInfo 2011: Innovations in Sharing Environmental Observations and Information
Copyright 2011 Shaker Verlag Aachen, ISBN: 978-3-8440-0451-9

Approaches to the Dynamic Energy Footprinting of Online Media[1]

Daniel Schien[2], Dr. Chris Preist[3], Paul Shabajee[4], Dr. Mike Yearworth[5]

Abstract

Energy consumption of digital media services in the Internet has previously been assessed based on assumptions of average use patterns and power draw of devices. With increasing popularity of these services also grows the need for models which accurately assess the power consumption for their provision. We present a dynamic modelling approach based on real-time measurements at datacentre servers, the network infrastructure and the end user devices which can serve to corroborate previous results and in the future design of more energy efficient digital services.

1. Introduction

Existing work on end-to-end energy and carbon life cycle analysis of digital services (Malmodin, Moberg, Lundén, Finnveden, & Lövehagen, 2010), (Chandaria, Hunter, & Williams, 2011), (Daniel R. Williams, 2011), (Baliga, Ayre, Hinton, Sorin, & Tucker, 2009), (Feldmann, Lange, & Kind, 2010) uses static modelling together with averaged, aggregate data regarding usage and power. However, to get a more accurate and nuanced value for the energy used by a particular instance of a digital service across the network as it is delivered requires integrating an LCA model of a digital service with real-time data collection regarding the resource usage of that service at a given moment. Different service customers will access the service from different locations, at different times, with different preferences, and under different loads on both the network and the datacentres involved in the delivery of the service. In this paper, we give an overview of such an approach, using as an example the delivery of a specific view of a multimedia news and entertainment website to a specific customer.

Having a more precise, dynamically generated assessment of the specific delivery of a webpage viewing (or other digital service) can be used in ways in which the standard aggregate approach cannot be. Firstly, it can be used to provide feedback to an online user as to the energy impacts of their online behaviour, with the potential for encouraging behaviour change. As (Preist & Shabajee, 2010) observe, it is not likely that this will have a significant impact on behaviour, particularly under the current network charging models, nor is it clear that this is necessarily a desirable approach. Secondly, and more importantly, it can be used (together with user behaviour data) to provide feedback to website designers as to the energy impact of specific design decisions on specific pages, and whether the energy 'cost' of features is merited given the observed 'value' to the user. Thirdly, while it may not be appropriate to dynamically assess all services, such an approach can be used to test assumptions made in more static assessment methodologies and either corroborate or challenge their validity.

The remainder of the paper is structured as follows. Section 2 will introduce the goal and scope of the assessment using the terminology of a life-cycle-assessment. Section 3 describes the model parts developed

---

[1] Research supported in the UK by EPSRC grant EP/I000151/1

[2] daniel.schien@bristol.ac.uk, Department of Computer Science, Faculty of Engineering, University of Bristol, BS8 1TR. UK

[3] chris.preist@bristol.ac.uk, Address as above

[4] paul.shabajee@bristol.ac.uk, Address as above

[5] mike.yearworth@bristol.ac.uk, Systems Centre, Faculty of Engineering, University of Bristol, BS8 1TR. UK

for the individual parts of the delivery system. Finally, in section 4 provides a discussion of the model and presents further work.

2. Goal and Scope

LCA methodologies (European Commission - JRC - Institute for Environment and Sustainability, 2011) require specific LCAs to clearly present their *goal* and *scope*. While what we present here is not a specific and rigorous LCA, we adopt this approach to provide clarity.

The goal of the methodology and system we are presenting is to allow the energy and carbon footprinting of a specific media webpage access, at a specific time from a specific location. This is intended to be used in the longer term to enable decision making with regard to both the design of the web site, and of the underlying architecture supporting it (both in the datacentre and over the broader internet).

The scope of an LCA includes several aspects. Firstly, the type of LCA to be adopted; an LCA can be *attributional,* allocating all emissions in the system under consideration on a pro-rata basis to appropriate system outputs. Alternatively, it can be *consequential*, considering the change in emissions from the system under consideration resulting from a change in product consumption - for example, if website usage rate was to increase by 10%. In line with past work in this area, we adopt an attributional approach for the purposes of this paper. We focus specifically on energy use, and do not consider other impact categories which are included in a complete LCA.

Secondly, the scope of the LCA requires a decision around the *functional unit* selected for analysis. The functional unit is a description of the product or service which is delivered from the system under consideration, and its characteristics. For example, a non-dynamic LCA of web page delivery could specify such features as the typical size of the web page, the number of javascript calls associated with it, the location of the typical viewer etc. However, in our case we determine the functional unit of the analysis to be *delivery of a specific web page to a specific user* as determined at the time of execution. Though we focus on this functional unit for the purposes of this paper, it is straightforward to see that this information can be aggregated to support other functional units likely to be of interest to a service provider - such as the footprint of a particular customer visit to the site, or the footprint of a given page on the site in a given day.

Finally in determining the scope, we need to consider the system and its boundaries. The system we are considering consists of the *energy use* involved in the delivery of a given *web page instance* from a service provider to a given customer, and viewing of that page. Hence the system under consideration includes the specific server activity involved in dynamically constructing and populating of the web page instance with content for the user. This will involve several servers systems such as web servers, application servers or database servers at the host site, but is also likely to involve server activity elsewhere - for example, in content delivery networks to accelerate the delivery of the site, and third party advertising servers selecting and delivering appropriate adverts for the page and user. When browsing an online media web page such as a newspaper or a video portal, users will generally receive the basic HTML document from the original service provider, but it in turn contains links referring to a server operated by the CDN. When a client browser attempts to fetch a linked resource, such as an audio, video or image, the CDN directs its request to the server likely to be able to provide the resource most quickly. As a consequence, individual users receive the data which makes up a given web page from different physical machines in different geographical locations and along routes of different number of hops and of different router models. We also consider other data centre energy usage, such as by peripherals and air conditioning equipment.

Beyond the data centre, we include the energy used by equipment used in both core and edge Internet networks in communication between client, host data centre and third party data centres involved. We also include network access devices of different types, mobile networks, home networks and client devices of different types.

The analysis does not include the embodied energy or the impact of the end-of-life phase of the devices involved, and does not include energy used during the generation of website content (e.g. by journalists, copy editors etc.) However, we do allow for the use of a development system running alongside the live system in the data centre, where new pages are uploaded and tested prior to going live.

3. Assessment techniques

In the previous section, we described the system under consideration in our analysis. We now consider each contributor to energy consumption within the system – provider data centre, network, third party data centre and end user. In each case, we describe our approach to modelling its energy consumption associated with the delivery of a given service, and describe what techniques we use to gather appropriate data in real-time. Where real-time primary data is not accessible, we describe what secondary sources we use as estimates.

3.1 Data Centre

To allocate energy use within the service provider datacentre to a given service instance, such as a web page request by a specific user, we need to measure what resources within the datacentre are used to serve this request, and then allocate energy usage from those resources on a pro-rata basis. Specifically, the energy using resources in a data centre are the servers, associated peripherals such as storage equipment, network equipment and power transformation as well as heating, ventilation and air conditioning (HVAC)equipment to service the data centre as a whole.

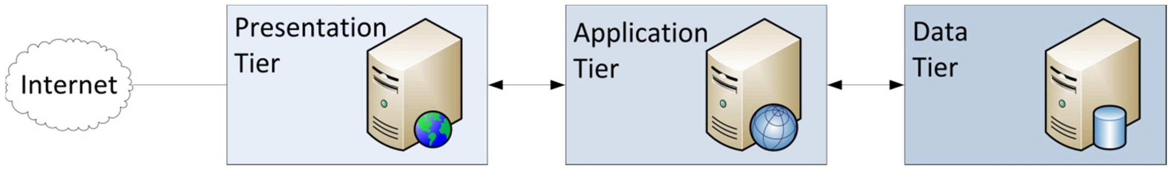

Figure 1 Multi-Tiered Datacentre Application

In delivering the multimedia news web service, we assume the provider adopts the commonplace multi-tier application architecture as illustrated in Figure 1 where concerns of the service delivery are separated between systems which are typically deployed as a number of virtual machines. Web pages are composed in response to a user request. An increasing part of web pages are customised for each user. This might include local weather forecast or customised advertisement. By combining power monitoring of individual devices and performance monitoring of process activity on these devices, it is possible to determine energy associated with a given service request. Firstly, we allocate energy usage to each virtual machine by using the resource utilisation model of (Rivoire & Ranganathan, 2008). In our initial model each virtual machine is responsible for a single function within the tiered application stack, and at any given time interval handles a number of service requests. We assume that, within a given virtual machine, each request uses roughly equivalent computational resources and hence apportion the energy used in that time interval evenly between the requests serviced. However, a given customer service will require different numbers of requests to different applications within the stack which can be inferred from application log files. In order to calculate the power per request power monitors measure the power draw on physical servers in small time intervals (t_i,t_j). Let M be the set of virtual machines which are involved in servicing the particular request. During this time, each system S_i in M services n_{S_i} requests. Not all machines in the datacentre have to be involved in servicing the particular request. Let n refer to the total number of requests served in the time interval. Based on (Daniel R. Williams, 2011) we allocate an energy overhead from the network devices in the datacentre and equally apportion it to all service requests during a given time interval n, and use a fixed value for the power draw of these devices because their energy consumption does not fluctuate with load (Hlavacs, Da Costa, & Pierson, 2009). We also add an additional coefficient for HVAC and power transformation, commonly referred to as power utilisation efficiency (PUE). Given these definitions the equation for the energy allocation to a given service request is:

$$E_M = (\sum_{S_i \in M} \frac{P_{S_i}}{n_{S_i}} + \frac{P_{NET}}{n}) \cdot PUE \cdot (t_j - t_i)$$

3.2 Network

The internet consists of *core*, *edge* and *access* networks. A number of core networks, high speed and high bandwidth, are operated semi-independently of each other and are referred to as *autonomous systems (AS).* These core networks are linked to each other, and to end access networks, via the edge networks. In turn, the access networks consist of home and office LAN (wireless and cabled), and mobile networks.

To allocate energy used in the network by specific service requests requires identifying which parts of the network have been used to service a given request, identifying the network devices involved, and allocating the energy proportionally according to what percentage of the traffic the specific service request was responsible for.

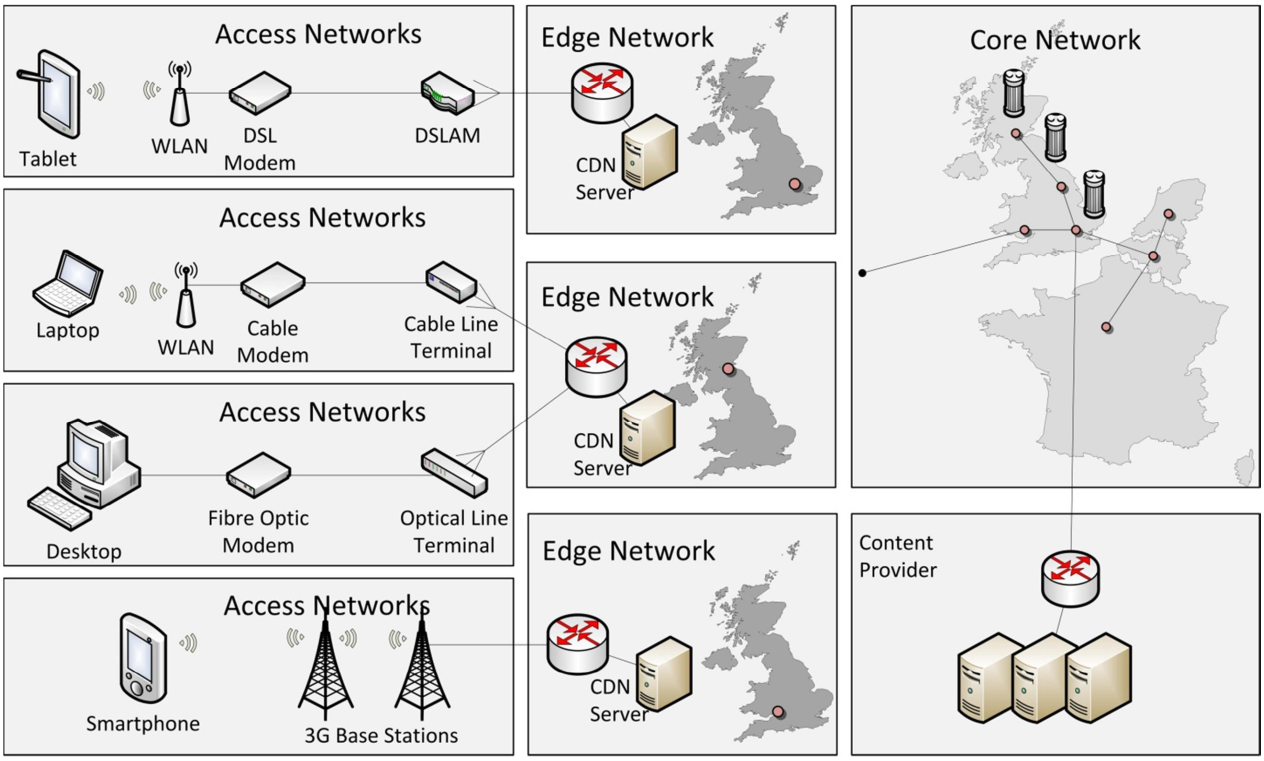

Figure 2 Network Model

The route taken through the network can be identified using the `traceroute` utility. This tool can identify network devices involved in a connection between provider and customer which provide routing services using the Internet Protocol[6]. However, it does not give information about which specific models the switches and routers involved are. To do this, we determine whether they are likely to be core or edge units, and then use power values for typical products identified from existing analyses (Baliga et al., 2009), (Lee, Rimac, Kilper, & Hilt, 2011), (Feldmann et al., 2010). To determine if a unit is likely to be core or edge, we use the following heuristic: The `traceroute` IP address allows identification of which autonomous system (AS) the unit belongs to. We assume that in any AS path of length four or greater all but two hops are core routers and that there are no core hops in ASs of length three or less. This assumption is based on similar descriptions of the mapping between the physical and the logical topology in networks found, among others, in (Medhi, 2007). Given this, we calculate the energy used by edge and core network devices by using equations based on those of (Baliga et al., 2009).

In addition to the network switches and routers identified by the `traceroute`, there is also additional energy consumption from internet transport systems, as traffic is forwarded both overland and through undersea cables. To determine the distance a signal must travel both overland and undersea, we use an IP geolocation service to identify the location of the end user and intermediate points on the path, together with data on the location of undersea cable landing points. Again, using variants on equations and data from (Baliga et al., 2009), we can estimate the energy involved in transportation for this specific service.

---

[6] In the OSI model the Internet Protocol (IP) is identified with layer 3 or the network layer. In edge and core networks IP enabled devices are connected by fiber optic links with higher capacity than Ethernet cables and which do not use IP but a set of protocols which are placed at lower layers of the OSI model.

A selection of the variety of common access network technologies which subscribers can choose from is illustrated in Figure 2. Line terminals provide a connection between a specific end user access technology such as DSL to the part of the Internet in which data is routed by the Internet Protocol. These devices such as DSLAMs, cable line terminals or the mobile base stations serve several end users simultaneously but the number of users who connect through these devices to the Internet is smaller than those who of devices in the edge or core network. We therefore assume that traffic in the access networks is less aggregated and throughput more variable. In other words, patterns of utilisation of access network devices show a higher variability as those of the edge and core network. At the same time, the power consumption of network devices is independent of their load (Hlavacs et al., 2009). We account for these two features by allocating the power consumption in the access network by the users' viewing time of the web page as opposed to the data volume transferred. Unfortunately, Internet service providers do not publish detailed information about the structure and subscription rate of their access networks. We therefore model the power consumption based on available average data.

3.3 Content Delivery Networks and other third parties.

As touched on in section 2, not all content the end user receives is sourced from the data centre of the service provider. Content delivery networks, ad networks and third party service providers (such as local weather) can all contribute data within a web page. In this discussion, we focus on Content Delivery Networks, though the principals can be applied to most other third party providers.

Content Delivery Networks (CDNs) such as those provided by Akamai[7] and Level 3[8] play an important role globally in increasing speed of end user access and reducing congestion by hosting third party content at distributed locations around the internet. The energy used in the process of interacting with a CDN consists of the energy of transmitting the data across the network, and the energy used at the CDN's datacentre. To calculate the energy used by the network, we have two options depending on if the end-user is an active participant in the analysis. If this is the case, a `traceroute` can be run from the end-user location to give accurate information about the network route, and the approach described in section 3.2 can be applied. If this is not possible, an estimate of a typical CDN route from (Huang, Wang, Li, & Ross, 2008) can be used instead. To calculate energy used in the CDN data centre, we adopt a modified version of the approach taken by (Chandaria et al., 2011) and apportion the energy used by a canonical CDN server according to proportion of its average load taken by a given service. In both cases, data is required on the size of the resources sent from the CDN. This is available from the service provider, who holds the original versions of each resource.

3.4 End User

The end user equipment includes an access device with a variable power consumption ranging from desktop PCs to mobile phones and optionally network devices for Internet access such as a wireless router. The power consumption of the access device depends on the resource utilisation of the applications executed on it. The power consumption of the network device is independent from utilisation (Hlavacs et al., 2009). In previous studies the power consumption of end users was estimated on average values without adjusting for the large variation of power consumption of individual devices. In order to improve on these estimations we require for the user to participate by providing information to the model. The number and kind

---

[7] http://www.akamai.com

[8] http://www.level3.com

of applications on the access device varies over time. For the web browser application variability of resource use results from the varying throughput of data as well as the variation of the type of media displayed. As an example, video decoding utilises more resources than rendering of text. In order to capture this variability we dynamically calculate power consumption for each application based on several performance indicators such as CPU, memory and I/O utilisation. We distinguish between user applications which the user interacts with, such as the browser, a word processing application or a music player and the system processes which support these applications. The total power of the end user device is the sum of the power consumption of all applications and processes. The system processes are shared between all processes and their resource utilisation is allocated to a specific application pro rata of user attention time to that application. We assume that the user at any single time can only lend his attention to a maximum of two separate applications one serving audible the other visual content. At this point of the modelling we assume for applications serving textual content the attention to be with the application which has the window focus. As with other simplifying assumptions we explore their consequences and options for refinement during the implementation of the models. Applications serving audio content are considered to constantly have the user attention. Then the base load of system processes is equally apportioned to the applications which have the user attention. The remaining base power consumption during idle phases during which no application has the user attention is allocated equally to all running applications. The resulting energy consumption of the end user device is the integrated power consumption of the web browser over the time of reading the web site. The power consumption for a specific degree of resource utilisation is device specific. Therefore, the model needs to be calibrated for every specific device configuration.

Optional network access devices such as wireless routers are usually not powered down when not in use and draw constant power independent of their data throughput. As a consequence, the power consumption of home networking in general is independent of the particular applications that users are executing. At this stage we allocate power for home networking based on the time of the page visit. Initially, average values for network device power consumption will be used. If user input is provided to the power model it can be improved with the specific power consumption of the particular networking model.

4. Discussion and Further Work

We have presented approaches to a dynamic energy footprinting of digital media. The technology for digital media delivery in the Internet is under constant development towards higher diversification of processes. Power models have to keep up with this trend and capture the dynamics of power consumption. While the existing end-to-end models provide conservative lower bound estimations (Baliga et al., 2009), (Feldmann et al., 2010) the models based on market size provide lose upper bounds (Koomey, 2007), (Malmodin et al., 2010). Our dynamic model improves on both of these approaches, particularly in regards of power consumption in the data centre and the end user.

As network operators do not disclose the power consumption of either their network devices or the traffic patterns in their networks for commercial reasons the existing modelling approaches are based on best practice assumptions. We build our model based on these assumptions and expand them by dynamically measuring the route length. Similar properties apply to the content delivery networks which assume an increasingly important role in the digital media distribution of the Internet.

The dynamic model does not directly provide an energy footprint for the entire sector. Instead we measure the network routes from the vantage point of a particular content provider. At the same time we attempt to build the most accurate energy model of the end user devices. This requires their active participation. The advantage of these resource utilisation based models is that they provide high accuracy (Rivoire & Ranganathan, 2008). However, we note that these models require calibration for each particular end user device and hence will only be available for standard end user devices with a limited number of hardware configurations.

We are currently in the process of implementing this model. We have developed the software models for the data centre and plan to trial them with a major news provider. We also have developed parsers for a growing number of publicly available `traceroute` servers to measure hop distances between system elements. A possible extension of the model is the inclusion of carbon emission data of electrical energy for each system part based on its geolocation.

Bibliography

Baliga, J., Ayre, R., Hinton, K., Sorin, W. V., & Tucker, R. S. (2009). Energy Consumption in Optical IP Networks. *Journal of Lightwave Technology, 27*(13), 2391-2403. doi:10.1109/JLT.2008.2010142

Chandaria, J., Hunter, J., & Williams, A. (2011). The carbon footprint of watching television, comparing digital terrestrial television with video-on-demand. *Sustainable Systems and Technology (ISSST), 2011 IEEE International Symposium on* (pp. 1–6). IEEE.

Daniel R. Williams, Y. T. (2011). A Methodology to Model the Energy and Greenhouse Gas Emissions of Electronic Software Distributions. Reading (Working Paper).

European Commission - JRC - Institute for Environment and Sustainability. (2011). ILCD Handbook. Retrieved July 28, 2011, from http://lct.jrc.ec.europa.eu/assessment/publications

Feldmann, A., Lange, C., & Kind, M. (2010). Energy Trade-offs among Content Delivery Architectures. *6th Conference on Telecommunication Techno-Economics, 2007. CTTE 2007*. Ghent. doi:10.1109/CTTE.2010.5557700

Hlavacs, H., Da Costa, G., & Pierson, J.-M. (2009). Energy Consumption of Residential and Professional Switches. *2009 International Conference on Computational Science and Engineering*, 240-246. Ieee. doi:10.1109/CSE.2009.244

Huang, C., Wang, A., Li, J., & Ross, K. W. (2008). Understanding hybrid CDN-P2P: why limelight needs its own Red Swoosh. *Proceedings of the 18th International Workshop on Network and Operating Systems Support for Digital Audio and Video* (pp. 75–80). ACM.

Koomey, J. G. (2007). *Estimating Total Power Consumption by Servers in the U.S. and the World*.

Lee, U., Rimac, I., Kilper, D., & Hilt, V. (2011). Toward energy-efficient content dissemination. *Network, IEEE*, *25*(2), 14–19. IEEE.

Malmodin, J., Moberg, Å., Lundén, D., Finnveden, G., & Lövehagen, N. (2010). Greenhouse Gas Emissions and Operational Electricity Use in the ICT and Entertainment & Media Sectors. *Journal of Industrial Ecology, 14*(5). doi:10.1111/j.1530-9290.2010.00278.x

Medhi, D. (2007). *Network routing: algorithms, protocols, and architectures*. *Computer Networks*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Preist, C., & Shabajee, P. (2010). Energy Use in the Media Cloud. *Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on*.

Rivoire, S., & Ranganathan, P. (2008). A comparison of high-level full-system power models. *2008 conference on Power*.