HomeFlow: Inferring Device Usage with Network Traces

Oliver Bates
School of Computing and Communications,
Lancaster University, UK
o.bates@lancaster.ac.uk

Matthew Broadbent
School of Computing and Communications,
Lancaster University, UK
m.broadbent@lancaster.ac.uk

UbiComp'13 Adjunct, September 8–12, 2013, Zurich, Switzerland.

Abstract
Previous studies in home energy have taken a service oriented approach to disaggregating direct energy consumption. With a particular focus on media and ICT services in the home, our proposed platform builds upon this work by providing activity oriented data, collected through home network monitoring. This information will be used to build a profile of communication between devices. This includes inter-device communication within the confines of a home environment, and also the use of external resources outside of the home. This provides knowledge of device behaviour and enables profiling of device relationships. Furthermore, monitoring communication to locations outside of the home will enable us to estimate associated indirect energy costs. These are incurred when a user consumes an externally provided service, such as Video-on-Demand.

Author Keywords
Home Networking, Home Energy, Micro accounting, Energy in practices

ACM Classification Keywords
H.5.m [Information interfaces and presentation]: Miscellaneous.

Introduction

The motivation behind this work is to further our understanding of how energy is consumed within the home. Bates et al [1] discuss how energy consumption is attributed to individual services within 4 student flats. They highlight that media and ICT services contribute between 3.5–34% of the total flat consumption.

HomeFlow aims to help us understand this large level of variance in media and ICT consumption. By monitoring the fine-grained interactions of networked devices, we can attribute usage based upon observed network traffic for these devices. It is important to note that this only improves our accuracy, it does not complete it. For example, a media device may not be networked at all, such as an audio amplifier. Similarly, multi-use devices may be used for an activity that does not require a network connection, such as word processing. Nonetheless, we still expect HomeFlow to offer a greater insight into home energy usage. As we move towards a more connected home environment, HomeFlow becomes even more relevant.

Bates et al. observed that their participants used media and ICT devices in conjunction with other similar devices to support particular activities and practices. This is referred to as "A system of devices". In this paper we will refer to these groupings of devices as constellations. HomeFlow will log and profile network data between devices and categorise network enabled devices into constellations.

Unlike previous work on single-point energy disaggregation [5] or per-socket monitoring [1] we are only concerned with collecting data that allows us to tie activities to media and ICT devices and services - we do not focus on direct energy implications.

A recent study by Kawsar et al. [6] uses a combination of network data, surveys and semi-structured interviews to tie performed activities back to devices in the home. Our proposed work draws on this, but concentrates on profiling relationships between devices in the home and other external services.

Chetty et al. [3] calls for a rethink in "applications and devices to accommodate those who experience the Internet via usage-based pricing'. ProjectBISMark [9] works towards this by monitoring external connectivity. HomeFlow indirectly supports both of these efforts: we bring much of this functionality together to create a novel data aggregation layer that can inform further energy decisions in a home environment. This focuses on creating constellations of devices using inter-device communication to infer relationships.

Mike Berners-Lee writes about the embodied carbon costs associated with datacentres. Datacentres contributed 130 million tonnes of CO2e in 2010, this is set to rise to 250–340 million tonnes CO2e by 2020 [2, p. 161 - 162]. A New York Times article discusses the negative energy impacts of powering and maintaining datacentres [7]. HomeFlow aims to store data that relates to these indirect energy costs by capturing and profiling traffic that leaves and enters the home network.

The aims of this platform are 1) to gain better understanding of the activities that are being performed by media and ICT devices through the monitoring of network traffic (e.g. are they watching NetFlix whilst emailing, or are they just playing Angry Birds?), 2) provide better understanding of constellations in the home by linking devices through network communication patterns, and 3) collection of external

traffic data that will allow for the attribution of indirect energy to devices and services within the home (e.g. streaming video-on-demand).

Proposed Method

HomeFlow is currently in early development stages and requires two elements for data collection and profiling. Our proposed platform includes a custom router running OpenWrt and a logging PC. The logging PC will also be used for the profiling of traffic.

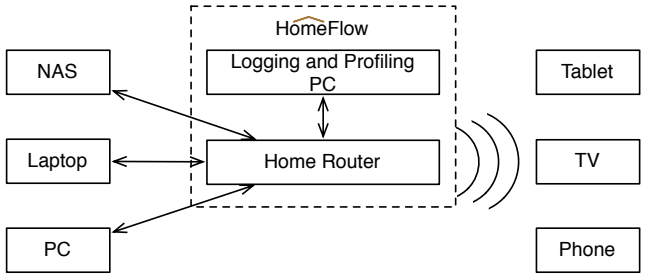

Figure 1: The proposed framework.

The HomeFlow platform (Figure 1.) will be used to collect and profile network data in the home. Initially, this will be categorised into two different groups: internal traffic and external traffic. We will use the industry-standard NetFlow format[1] to store this information. NetFlow was chosen because it has a relatively lightweight collection process, and can be run on a resource constrained device, such as a home gateway. It facilitates the collection of traffic statistics into summarised flows, which reduces storage requirements and permits faster parsing. A NetFlow exporter is also an installable package on OpenWRT[2] firmware, which affords us greater control of the router.

We can characterise communication by examining the source and destination IP addresses of recently captured data. Not only does this enable us to determine which device is communicating with which, it also allows us to monitor the characteristics of such flows: protocols and ports used, packets and bytes sent, and the Type of Service fields set. We can also filter irrelevant information which does not constitute a noteworthy interaction between devices. Furthermore, HomeFlow allows us to determine the *nature* of the interaction: is it video streaming? is it gaming traffic? is it a cloud backup?

With this information, we can also infer device constellations and connections - when a particular device is turned on, this other device is also used in conjunction. Furthermore, we can use these relationship inversely: when one device is turned on, this other device can be turned off.

These inferences are not limited to device connections either. Using NetFlow data, we can also investigate the times and durations of usage. Once we have built profiles of these patterns, we can identify when a device is acting outside of its normal behaviour. This is particularly pertinent if we can use this information to turn off always-on devices, and prevent them from making external interactions during inactive periods. For example, we can turn a network interface off on a games console in the early hours of the morning to stop it from contacting cloud services and performing an update. We argue that such a process can be

[1] NetFlow - **www.cisco.com/go/netflow**

[2] OpenWRT - **https://openwrt.org/**

completed during normal operation without keeping the device always-on.

Privacy
Data collected from the HomeFlow platform contains personal and private information. Ensuring that this data is kept secure and used properly is of upmost importance. To this end, HomeFlow keeps the processing of information within the confines of a deployment: data is processed within this environment and is not exported externally under normal conditions. It is envisaged that results will only be available to the members of the household through a secure interface requiring authentication.

If data is to exported, particularly for the use in research and during development, we will ensure that participants are fully aware of the nature of this work and their full consent will be sought. Alternatively, network data can be anonymised to avoid results being directly linked back to specific users. In the case of HomeFlow, this would only be possibly on source IP addresses; destination IP addresses are a pivotal part of inferring service usage, and as such must be left intact to remain useful.

Deploying in the Home

HomeFlow will be deployed in homes to collect network data and allow for the profiling of devices and constellations. The data will be collected per device over time and profiled. Automated constellation constellation learning will occur through the observation of the profiled device traffic.

HomeFlow was originally intended as a tool for researchers to understand media and ICT device constellations and communication. If we were to design a visualisation component for HomeFlow it would visualise live and historical traffic for devices and constellations. We would associate chunks of device and constellation traffic with activities. Using similar methods to Costanza et al. [4] we would allow users to annotate these chunks with a specific activity or by tagging users. By doing this we would gain a better idea of the activities supported and who is using the devices and constellations over time.

The visualisation will be available through a local web portal within the home, ran from the logging and profiling machine. The proposed visualisation component will be web-based, and therefore it will be possible to display this information on both existing home automation components with displays and/or users own Internet enabled devices.

The only additional direct energy costs of this deployment will be incurred by the logging/profiling machine. The router used to collect data will be a typical home router, with the slight modification of running a NetFlow collector.

Combining with energy data
If combined with a data from single-point or per-socket energy monitors, energy usage could also be attributed to particular devices, constellations, activities or users. This data could also be incorporated in the visualisation, for example displaying chunks of bandwidth and energy consumption for a given constellation over time.

Implications for Ubicomp

Accountability of service impact. Combining profiled HomeFlow data with per-socket energy data enables us to closely tie direct energy consumption to activities

performed by media and ICT devices. Furthermore, by incorporating our collected external network data with 1) methods of accounting for indirect energy costs of cloud computing tasks [10] and 2) location impacts on the energy footprint of digital media [8], we can build a better overall representation of the energy impacts our media and ICT constellations carry.

Recognising energy waste HomeFlow enables profiling of internal and external traffic outside of the normally perceived operating hours of devices. With the incorporation of per-device energy consumption data, we will be able to identify such out-of-hours communication and move to reduce it. For example, if an audio receiver is powered up at 4am this could be profiled as 'waste energy' if 1) it is the only device consuming power as part of a constellation, or 2) it is not receiving any audio streams over the network.

Supporting smarter homes. By feeding profiled hours of waste energy from HomeFlow into a home automation system would support the configuration of automated tasks for the switching off of out of hours and wasteful devices. For example, you can use constellation data to infer relationships: if the TV is off, then the games console can be too.

Challenges going forward Currently, non-networked devices will not be captured using HomeFlow, e.g. an audio receiver using optical output from a TV or Blu-ray player. Tracing traffic to external services, especially those in the cloud, is non-trivial. Most modern services are distributed in nature, and as a result, service address can resolve to a number of network locations. In addition, we must also consider background infrastructure and services that have a requirement to always be on, such as: wireless routers, home servers, and home automation systems themselves.

Acknowledgements

We gratefully acknowledge the support of our funders Microsoft Research and the Faculty of Science and Technology at Lancaster University.

References

[1] Bates, O., Clear, A. K., Friday, A., Hazas, M., and Morley, J. Accounting for energy-reliant services within everyday life at home. In *Proc. of Pervasive* (June 2012).

[2] Berners-Lee, M. *How Bad Are Bananas?: The carbon footprint of everything*. Profile, 2010.

[3] Chetty, M., Banks, R., Brush, A., Donner, J., and Grinter, R. You're capped: understanding the effects of bandwidth caps on broadband use in the home. In *Proc. of CHI*, CHI '12, ACM (New York, NY, USA, 2012), 3021–3030.

[4] Costanza, E., Ramchurn, S. D., and Jennings, N. R. Understanding domestic energy consumption through interactive visualisation: a field study. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, ACM (New York, NY, USA, 2012), 216–225.

[5] Froehlich, J., Larson, E., Gupta, S., Cohn, G., Reynolds, M., and Patel, S. Disaggregated end-use energy sensing for the smart grid. *Pervasive Computing, IEEE 10*, 1 (2011).

[6] Kawsar, F., and Brush, A. Home computing unplugged: Why, where and when people use different devices at home. In *Proc. of Ubicomp 2013* (2013).

[7] New York Times. Data Centers in Rural Washington State Gobble Power. http://nyti.ms/12sA5P1, 2012.

[8] Schien, D., Preist, C., Yearworth, M., and Shabajee, P. Impact of location on the energy footprint of digital media. In *Sustainable Systems and Technology (ISSST), 2012 IEEE International Symposium on* (2012), 1–6.

[9] Sundaresan, S., de Donato, W., Feamster, N., Teixeira, R., Crawford, S., and Pescapè, A. Broadband internet performance: a view from the gateway. *SIGCOMM Comput. Commun. Rev. 41*, 4 (Aug. 2011), 134–145.

[10] Teehan, P., Kandlikar, M., and Dowlatabadi, H. Estimating the changing environmental impacts of ict-based tasks: A top-down approach. In *Sustainable Systems and Technology (ISSST), 2010 IEEE International Symposium on* (2010).