Trying broadband characterization at home

Mario A. Sánchez, John S. Otto,
Zachary S. Bischof, Fabián E. Bustamante
{msanchez, jotto, zbischof, fabianb}@eecs.northwestern.edu

Northwestern University

Abstract. In recent years the quantity and diversity of Internet-enabled consumer devices in the home have increased significantly. These trends complicate device usability and home resource management and have implications for crowdsourced approaches to broadband characterization. The UPnP protocol has emerged as an open standard for device and service discovery to simplify device usability and resource management in home networks. In this work, we leverage UPnP to understand the dynamics of home device usage, both at a macro and micro level, and to sketch an effective approach to broadband characterization that runs behind the last meter.
Using UPnP measurements collected from over 13K end users, we show that while home networks can be quite complex, the number of devices that actively and regularly connect to the Internet is limited. Furthermore, we find a high correlation between the number of UPnP-enabled devices in home networks and the presence of UPnP-enabled gateways, and show how this can be leveraged for effective broadband characterization.

1 Introduction

Over the last few years we have seen a dramatic increase in the quantity and diversity of Internet-enabled consumer devices in the home. Recent reports suggest that shipments of Internet-ready electronic devices – such as televisions and video game consoles – will surpass 500M units by 2013, triple the amount shipped in 2010. [1] This unparalleled growth challenges home network usability and resource management, and has implications for broadband characterization behind the last mile [1, 7, 8, 10].

The Universal Plug and Play (UPnP) protocol has emerged as an open standard to address some of these challenges [11], with a growing number of devices supporting it. [2] In this work, we leverage UPnP to understand the dynamics of home device usage and to sketch an effective approach to broadband characterization that runs behind the last meter. Previous studies have used

[1] http://www.isuppli.com/home-and-consumer-electronics/news/pages/shipments-of-internet-enabled-consumer-devices-to-exceed-pcs-in-2013.aspx
[2] http://realwire.com/releases/UPnP-Technology-Adoption-Continues-to-Soar-With-New-Areas-of-Growth

UPnP data to better understand home networks, focusing on characterizing the number and type of devices present [5] or inferring network characteristics including bandwidth and packet loss rates [4]. However, these studies have typically been based on single snapshot tests. In contrast, our analysis is based on measurements collected *continuously* from over 13K Dasu [9] users and studies the implications of this on broadband characterization.

We use the collected data to show the complexity of home networks in terms of number and type of devices detected (Sec. 3). We classify the devices found based on their likelihood of generating cross-traffic on the access link and analyze the dynamics of devices usage both at a macro (when devices are on/off) and micro level (when turned-on devices exchange data). We demonstrate that while in many cases the number of devices in the network is high, only a few of them actively and regularly connect to the Internet, potentially interfering with network measurements (Sec. 4). Furthermore, we find a strong correlation between the number of UPnP-enabled devices in the home network and the presence of UPnP-enabled gateways and suggest how this can be leveraged for effective broadband characterization from the home (Sec. 5).

2 Data Collection and Dataset

We conduct our analysis using data collected with Dasu, a platform aimed at broadband characterization and network experimentation [9]. We use a combination of passive and limited active measurements gathered over a 6-month period between February 24, 2012 and August 23, 2012. This dataset includes traces of BitTorrent and overall home network activity collected by Dasu from 13,605 homes spanning 151 countries.[3]

Each Dasu client periodically (at 30s intervals) collected anonymized traffic traces from BitTorrent's activity, including the number of bytes uploaded and downloaded as well as the current transfer speed, the total number of bytes sent/received was also captured using `netstat`. Beyond this passively collected data, clients also scanned the local network in search of Internet gateway devices using UPnP, following an approach based on DiCioccio et al. [5, 6].

For each gateway device responding to UPnP discovery messages, Dasu pulled their device definition XML data and collected the following configuration parameters: (a) current state of NAT for this connection, (b) external IP address, (c) current connection type (Cable, DSL), (d) maximum upstream/downstream bit rate available, (e) device model name and version. At the same rate, clients also retrieved dynamic information from the gateway including (f) cumulative count of bytes and packets received and (g) sent, as well as (h) the connection status.

A subset of clients periodically broadcasted UPnP discovery messages and recorded, for each responding device: (a) devices' uuid and UDN, (b) device type, (c) manufacturer, (d) model name and (e) model number.

[3] The dataset is available to other researchers upon request.

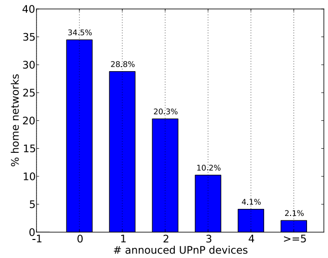


Fig. 1: UPnP-enabled devices in home networks.

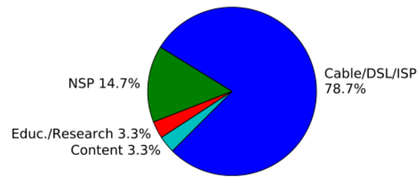


Fig. 2: Connection types for locations with ≥ 5 UPnP devices.

Since Dasu is implemented as a BitTorrent client extension, there is a possible concern that our conclusions could be affected by some type of bias common to BitTorrent users, such as a particular set of countries, connection or user type. We argue that BitTorrent users can be seen as early adopters and thus, in a sense, worst-case scenarios in terms of the level of complexity in home networks. In the following section we show that the collected dataset comes almost entirely (93%) from clients in typical residential networks and spread over a diverse set of nearly 100 countries.

3 The Home Network – A Complex Environment

In this section, we examine the complexity of home networks in terms of number and diversity of connected devices. Given our end-goal of deriving an effective approach to crowdsourced broadband characterization from end-hosts, we present our findings in this context.

We first look at the number of networked devices found, which we estimate for a subset of ≈4.6K of our client's locations using UPnP discovery messages. Figure 1 shows the distribution of clients' locations by the number of UPnP announced devices found. While 34.5% of sampled locations have no UPnP devices announcing their presence, over 65% of them has at least 1 device, and over 16% have 3 or more devices.

We know of two possible sources of errors in this estimation. By relying on UPnP discovery messages, our measurement approach can miss devices that do not support UPnP. On the other hand, it is also possible that multiple UPnP services can be hosted by the same device, so that by counting each announced-service as a different "device" we might be over-counting the number of UPnP-enabled devices in the network. We plan to address both issues as part of our future work.

To evaluate potential biases in our dataset, we analyze the distribution of sampled locations based on type of network connection. Type of connection can indicate something other than a residential network (such as educational or enterprise) which could bias our results, especially for locations with large number of UPnP devices. We focus thus our attention on those locations in

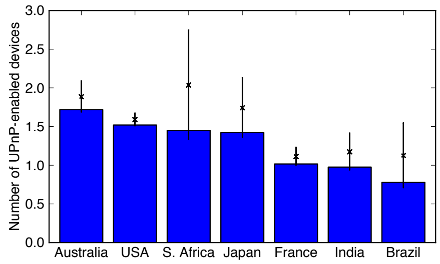


Fig. 3: Mean number of announced devices for homes across different countries.

Table 1: Top ten countries with > 1% of homes

Country	**%**	**Country**	**%**
Italy	2.68%	Switzerland	4.62%
Austria	2.95%	Germany	4.91%
Portugal	2.95%	Great Britain	6.42%
Canada	4.17%	France	8.03%
Australia	4.35%	USA	25.61%

our dataset with five or more devices. There are 96 such locations distributed over 61 different autonomous systems (ASes). We map most of these ASes to their business type using the peeringDb[4] database and manually label those for which we could not find an entry. Figure 2 shows the percentage of location per business type (i.e. *Network Service Provider*, *Education/Research*, *Content*, and *Cable/DSL/ISP*). As the figure shows, the sampled dataset comes almost entirely from broadband providers (i.e., Cable/DSL/ISP and NSP). Interestingly, the average (and median) numbers of devices for each of these business types are very similar ranging between 5 and 6.5 devices.

We now analyze the adoption of UPnP across different countries by looking at the number of UPnP-enabled devices connected to the sampled locations in different countries. For this analysis we restrict our set to countries with more than 50 homes and select for each home the snapshot with the largest number of announced devices across all samples. In terms of potential biases due to the countries where our clients are located (such as an unexpected fraction of locations in a few high-income countries), we find that the sampled locations come from ≈100 different countries, with ten or more locations in nearly half of them. Table 1 shows the top ten countries in our dataset with more than 1% of sample locations.

Figure 3 plots the mean number of announced devices for homes across different countries. The bars show the lower bound on the 1-sided 95% confidence interval, the line shows the 2-sided 95% confidence interval, and the X plots the mean value across all samples. We used the Student's t-distribution to compute the confidence intervals (as the population's standard deviation is unknown). The figure shows that high(er) income countries tend to have a higher number of UPnP-enabled devices in the home network.

To study the diversity of home network devices we classify the found UPnP-enabled devices and study their prevalence. For common devices we use the DLNA's "Home Network Device" specification[5] to categorize them and divide

[4] https://www.peeringdb.com
[5] http://dlna.org/dlna-for-industry/digital-living/how-it-works/dlna-device-classes

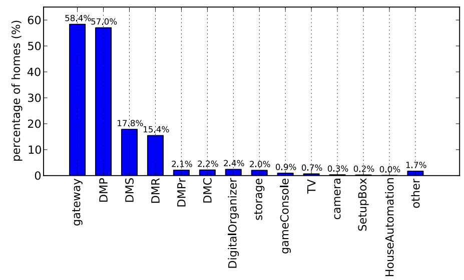


Fig. 4: The different UPnP devices and their popularity.

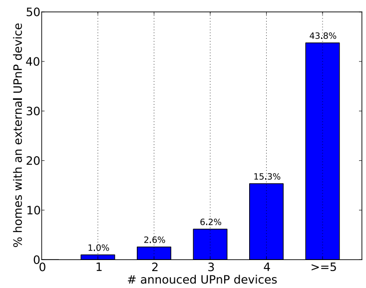


Fig. 5: Percentage of homes with external devices based on number of UPnP-announced devices.

the rest into functional classes such as *storage*, *cameras* or *television*. We labeled each device class as *Internal* and *External* based on their dominant network role – *externally-facing* devices that exchange traffic with the outside world (e.g. TV) or *internally-facing* devices that exchange traffic mostly within the home network (e.g. Storage). Given that the purpose of DLNA devices is to share media within the home (e.g. Digital Organizers, and Storage), each of these device classes are labeled as *Internal*. We classify the remaining classes of devices as external, including *Others*. We treat the *Gateway* category (e.g. DSL modems, WiFi routers) as its own class.

Table 2: Different classes of UPnP-enabled devices and their prevalence.

Device Type	**Connection**	**Perc.**
Gateway	Gateway	36.7%
Digital Media Player (DMP)	Internal	34.7%
Digital Media Server (DMS)	Internal	10.2%
Digital Media Renderer (DMR)	Internal	9.5%
Digital Media Printer (DMPr)	Internal	1.2%
Digital Media Controller (DMC)	Internal	1.2%
Digital Organizer	Internal	1.3%
Storage	Internal	1.1%
Game Console	External	0.5%
TV	External	0.3%
Camera	External	0.2%
SetupBox	External	0.1%
House Automation	External	$<$ 0.1%
Other	External	1.5%

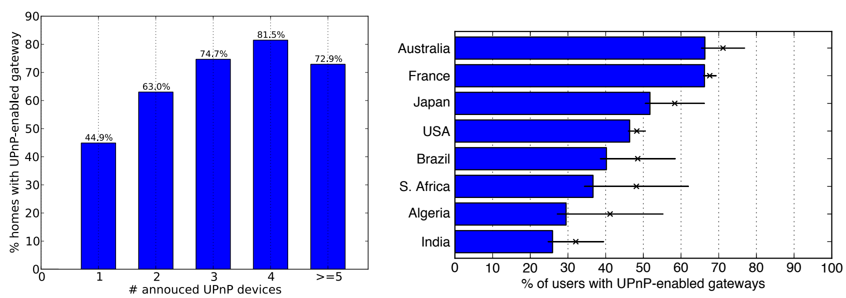
Table 2 shows the different device classes identified in our traces. From the ≈6K devices seen across ≈3K peers the most popular device type are gateways (over 35%) followed by a large number of DLNA-compliant devices, including Digital Media Players (34.7%), Digital Media Servers (10.2%) and Digital Media Renderers (9.5%). Changing focus to the distribution of these devices in the sampled locations, Fig. 4 plots the popularity of each device type across the studied locations with at least one UPnP-enabled device in their network. We note the high popularity of Digital Media Players, Servers and Renderers.

In the context of broadband characterization, we are particularly interested in the distribution of *internally-* and *externally-facing* devices. Figure 5 shows the fraction of home networks within each group for which at least one externally-facing device was identified. Not surprisingly, as the number of announced devices in the network increases so does the probability that at least one of those devices be an external device.

3.1 Prevalence of UPnP-enabled gateways

UPnP-enabled gateways are helpful for managing resources and monitoring the state of the network. Although UPnP-enabled gateways are not always available, their presence is particularly important in home networks with high number of devices, where cross-traffic could interfere with characterization.

Figure 6a shows the availability of UPnP-enabled home gateways in our sample. The figure plots the fraction of homes, with a given number of UPnP devices, in which such a gateway is present. As the number of UPnP-enabled devices in the local network increases, so does the likelihood that the home gateway supports UPnP.



(a) Homes (b) Countries

Fig. 6: Prevalence of UPnP-enabled gateways in sampled homes, clustered based on number of UPnP-enabled devices announced (6a) and (for a subset) by country (6b).

Last, we examine variations in the prevalence of UPnP-enabled gateways across countries. For this purpose, if a UPnP gateway is ever seen in a home network, we consider that sample as having a UPnP-enabled gateway. We group these homes by their ISP's country, giving us an estimate of each country's percentage of homes with UPnP-enabled gateways. We treat the data for each country as a sample from a binomial distribution and use the Wilson method to estimate confidence intervals.

Figure 6b plots the prevalence of UPnP-enabled gateways for several countries in our dataset. To account for different sample sizes across countries, we use the lower bound of the one-sided 95% confidence interval as a conservative estimate of the percentage of homes with UPnP in a given country. This means that there is a 95% chance that the actual percentage is *at least* the shown value. On the x axis, we show the proportion of samples having UPnP-enabled gateways and lines showing the extent of the two-sided 95% confidence intervals. In general, we find that more developed countries tend to have higher rates of UPnP-enabled gateways (as well as more complex home networks), hinting at a possible trend towards better environments for broadband characterization from end systems [2, 3].

4 Device Usage Dynamics

As the number of devices connected to the home network increases, so does the likelihood that the access link will be used by multiple devices simultaneously, potentially interfering with measurements looking to characterize the access link. In this section, we analyze the macro dynamics of network device usage – the frequency with which devices in the home network are active. We look at the micro dynamics – the rate and volume of traffic generated by these device – in the next section.

To study device dynamics, we leverage the fact that Dasu runs for long periods at a time (the median session time of a client is 178 minutes) and is thus able to take multiple snapshots of the active UPnP-enabled devices present on the network over time. We restrict the set of home networks to those for which we have at least 10 different sample snapshots and where there is more than one UPnP-enabled device announced and at least one of those is *outer-facing*. This set consists of 502 different home networks.

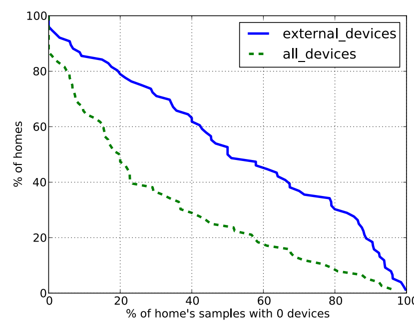


Fig. 7: Distribution of the fraction of homes vs the fraction of samples for which no other UPnP-device is present in the network.

We rank all locations based on the percentage of measurement samples where we find no other device/no external device active other than the host machine. Figure 7 plots the CDF for both – any device active (labeled *all_devices*) and external device active (labeled *external_device*). As the figure shows, for nearly 85% of the locations, the host computer where our measurement client is running is the only active external device in the network for at least 10% of measurement samples. For the median location, about 20% of the measurement samples occur when the host computer is the only active device in the network and nearly 50% of them when there is no other external device present.

5 Broadband Characterization with UPnP Help

The following two sections sketch an approach for effective end-system-based broadband characterization that takes advantage of UPnP-enabled gateways and illustrate its use with specific traffic scenarios.

Two sources of concern for broadband characterization from end systems are the presence of cross-traffic from other applications in the hosting devices and from other devices in the home network. We use `netstat`, a network statistics tool available in most platforms, to capture the number of bytes sent and received from the host and compare it against the amount of traffic monitored by our client. This allows us to identify situations where significant amount of traffic is being generated by other applications in the host device.

The second type of cross traffic is the one generated by other devices in the network. To identify such cases we employ the technique described by DiCioccio et al. [4] where UPnP-enabled home gateways are periodically queried to measure traffic in the home network. In cases where the UPnP-supplied data is both available and accurate, the authors showed that this technique provides a rich source of information for inferring the presence of cross traffic in the home network. Thus, for homes with UPnP-enabled gateways, we periodically query for traffic counters across its WAN interface (the number of bytes and packets sent and received). When we identify times where the number of packets or bytes sent or received is high enough to affect our measurements we simply discard (if passive) or postpone (if active) our measurements. While gateway UPnP traffic counters are not always accurate [4], such instances can be easily identified and accounted for.

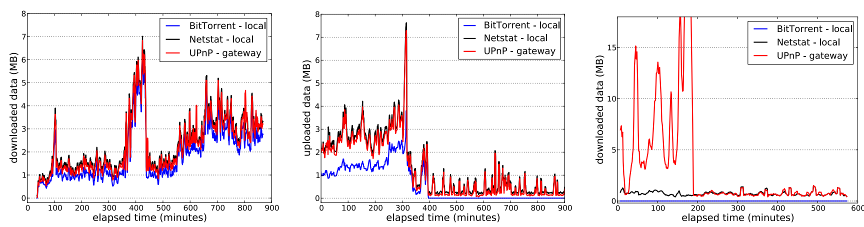
5.1 The Value of UPnP-Counters

We now present some concrete examples of how traffic counters from UPnP-enabled gateways allow us to disambiguate between different scenarios inside the home network. Using data collected from our Dasu users we show, for instance, how the presence of internal traffic can be identified and separated from traffic that uses the access link, both from the local host and other devices within the network.

No cross-traffic. As explained in Sec. 2, our traces contain the network activity as seen by each individual Dasu client at three different granularities. (*i*) Because Dasu runs as part of a network intensive application (BitTorrent) our traces contain traffic statistics about the number of bytes sent and received by the application alone. (*ii*) By using `netstat`, these traces also contain the overall traffic activity of the host, including the traffic generated simultaneously by all running applications at the time of collection. Finally, (*iii*) the client collects UPnP-supplied traffic data from the gateway which includes the number of bytes sent and received across the gateway's WAN interface.

Figure 8a shows the simplest scenario – where BitTorrent is solely responsible for the network traffic using the access link and the only source of traffic generated by the host. The figure plots the download activity of one Dasu client in a span of 15 hours in August 2012. Each of the three signals in the graph represents the number of downloaded bytes as reported by BitTorrent (blue), `netstat` (black), and the gateway counters (red), respectively, in intervals of 30 seconds increment. As the figure shows, all three signals overlap when Dasu' hosting application (BitTorrent) is the only network active application.

Local cross-traffic from other applications. Figure 8b plots the upload activity of another client, also for a span of 15 hours in June 2012. As before, the client is solely responsible for all the traffic present in the access link, but here BitTorrent is not the only network active application. As the figure shows, the signals that correspond to the local `netstat` counters (black) and the UPnP-counters at the gateway (red) overlap through the entire collection period (i.e., the client is the only device using the access link), but the signal that corresponds to BitTorrent traffic (blue) is much lower than that of `netstat` for the first five hours (300 minutes) of the session.



(a) No cross-traffic. (b) Local cross-traffic (up). (c) Cross-traffic (down).

Fig. 8: Traffic scenarios within the home network: (8a) download with no cross-traffic, (8b) local cross-traffic from other applications and (8c) download cross-traffic.

Cross-traffic from other devices. Figure 8c shows our last scenario, where there is significant cross-traffic from other devices in the home network. The figure plots download activity seen from a client over a span of five hours. In this case, there's no BitTorrent content being downloaded (the BitTorrent signal is a flat horizontal line around 0 bytes), but there is local traffic being generated by other applications in the host device (denoted by the black signal). However, for the first ≈ 200 minutes of the session, the traffic generated by the host devices represents only a small fraction of the total traffic present in the access link (red signal). The figure also shows the easily identifiable point at which the cross-traffic disappears.

6 Conclusion

The increasing complexity of home networks complicates device usability and home resource management and has implications for crowdsourced approaches to broadband characterization. In this work, we rely on UPnP measurements collected from over 13k end users study the complexity of home networks around the world. We presented a first look at the home network usage, both at a macro and micro level, and sketched an effective approach to broadband characterization that runs behind the last meter.

Acknowledgements

We would like to thank our shepherd, Aaditeshwar Seth, and the anonymous reviewers for their valuable feedback. We are always grateful to Paul Gardner for his assistance with Vuze and the users of our software for their invaluable data. This work was supported in part by the National Science Foundation through Awards CNS 1218287, CNS 0917233 and II 0855253 and by a generous Google Faculty Research Award.

References

1. Bischof, Z. S., Otto, J. S., and Bustamante, F. E. Up, down, and around the stack: ISP characterization from network intensive applications. In *Proc. of W-MUST* (2012).
2. Bischof, Z. S., Otto, J. S., Sánchez, M. A., Rula, J. P., Choffnes, D. R., and Bustamante, F. E. Crowdsourcing ISP characterization to the network edge. In *Proc. of W-MUST* (2011).
3. Canadi, I., Barford, P., and Sommers, J. Revisiting broadband performance. In *Proc. of IMC* (2012).
4. DiCioccio, L., Teixeira, R., May, M., and Kreibich, C. Probe and pray: Using UPnP for home network measurements. In *Proc. of PAM* (2012).
5. DiCioccio, L., Teixeira, R., and Rosenberg, C. Characterizing Home Networks With HomeNet Profiler. Tech. rep., Technicolor, 09 2011. CP-PRL-2011-09-0001.
6. DiCioccio, L., Teixeira, R., and Rosenberg, C. Measuring and characterizing home networks. In *Proc. of ACM SIGMETRICS* (2012).
7. Dischinger, M., Marcon, M., Guha, S., Gummadi, K. P., Mahajan, R., and Saroiu, S. Glasnost: enabling end users to detect traffic differentiation. In *Proc. of USENIX NSDI* (2010).
8. Kreibich, C., Weaver, N., Nechaev, B., and Paxson, V. Netalyzr: illuminating the edge network. In *Proc. of IMC* (2010).
9. Sánchez, M. A., Otto, J. S., Bischof, Z. S., Choffnes, D. R., Bustamante, F. E., Krishnamurthy, B., and Willinger, W. Dasu: Pushing experiments to the Internet's edge. In *Proc. of USENIX NSDI* (2013).
10. Sundaresan, S., de Donato, W., Feamster, N., Teixeira, R., Crawford, S., and Pescapè, A. Broadband Internet performance: a view from the gateway. In *Proc. of ACM SIGCOMM* (2011).
11. UPnP Forum. UPnP Device Management - Simplify the Administration of your Devices. Tech. rep., University of Zurich, Department of Informatics, 04 2011.