# Enhancing Visibility into Home Networks using SDN

Prithvi Sriram*, Veera Raghava Datla†, Hassan Habibi Gharakheili†, and Vijay Sivaraman†
*Department of Mathematics, Indian Institute of Technology, Guwahati, India.
†Electrical Engineering and Telecommunications, University of New South Wales, Sydney, Australia.
Emails: {prithvi@iitg.ac.in, v.datla@unsw.edu.au, h.habibi@unsw.edu.au, vijay@unsw.edu.au}

***Abstract*—The home network, typically shared by many household members and devices, is getting increasingly complex, yet neither ISPs nor subscribers have much visibility into aspects such as device connectivity patterns, user data consumption, and suitability of sites visited. In this paper we propose, develop, deploy, and evaluate a Software Defined Networking (SDN) based approach to enhancing visibility into home networks. We first develop an SDN architecture that leverages commodity residential gateway hardware and cloud-based software to provide real-time visibility into devices connected in the home, their data consumption patterns on an hourly and daily basis, and the domains visited by each device. We deploy our fully-functional system in selected households, and analyze their activity data collected over a month. Finally we analyze this data to present insights in terms of number and composition of connected devices, video-viewing patterns of data-intensive devices, and content preferences based on web-sites visited.**

## I. Introduction

Home networks are becoming increasingly complex, with many households having in excess of 10-20 Internet capable devices, including computers, tablets, smart-phones, smart-TVs, gaming consoles, household appliances, and medical devices. Further, several household users, using different devices to access different services, share the broadband connection. The subscriber lacks visibility into what is happening in their own network, such as which device is using how much data at any point in time, and how much of the monthly quota has been consumed by each device. Further, they have little control over which device is accessing what services, for example whether the kids are accessing inappropriate adult sites or engaging in social-networking during bed-time.

ISPs today avoid device-level visibility into the house, and typically deem their service to terminate at the external-facing port of the home gateway or router. This unburdens them from dealing with misbehavior or misconfiguration in (a heterogeneous set of) home devices (including the home gateway itself), which would otherwise consume excessive effort to support. From the ISPs point-of-view, value-add services for home users are simply not lucrative, particularly since the residential broadband market has low profit margins.

The above approach by ISPs leaves consumers high-and-dry, as they face new pain-points arising from the increasing number of household devices sharing the residential broadband link. Internet users in the developing world countries have a limited monthly quota on their Internet data volume consumption, of which an unfair share may get used by one or more household members/devices; these often lead to support calls by the consumer questioning how the data volume was used, which the ISP is unable to answer as they have no visibility of culpable device(s). Parents in the developed world are struggling to keep their kids safe from objectionable content over the Internet, but lack tools to do so.

One might be tempted to embed solutions for the above problems into home routers/gateways [1], [2]. We reject such an approach for multiple reasons: (a) home routers compete on price, limiting the software expertise available to vendors to embed advanced features; (b) consumers find embedded features hard to use [3]; and (c) embedded features get outdated rapidly. We therefore believe that the ISP is best positioned to serve the above needs, and Software Defined Networking (SDN) provides the necessary means to do so cost-effectively and at-scale, as argued next.

The decoupling of the data and control planes under the SDN paradigm allows an elegant approach to address the above needs: the home gateway forwards data packets using match-action rules compliant with the OpenFlow standard, while the control logic that implements the features above resides in the cloud. This simple architecture allows any commodity home router, enabled by an OpenFlow agent, to be used, bringing with it two significant advantages: (a) the ISP need not provide custom hardware or firmware, and can thus keep costs low by leveraging the competitive marketplace for home routers, and (b) new features can be enabled scalably from the cloud.

Our specific contributions are as follows. First, we build an SDN-based system architecture that enhances visibility into activity of connected devices in home networks, by tracking the presence of devices on the network, volume of data they use, and domains they visit. Our second contribution is to deploy our system in three households (authors of this paper) and analyze their network data corresponding to device-specific Internet activity of each household. Lastly, we apply some data analytics to infer insights about composition of households devices (i.e. personal, shared, or visitor), pattern of video consumption, and profile of preferred contents.

The rest of this paper is organized as follows: §II describes relevant prior work. We present our SDN system and infrastructure in §III. The analysis of household devices is in §IV, and in §V we present our insights and inferences. The paper is concluded in §VI.

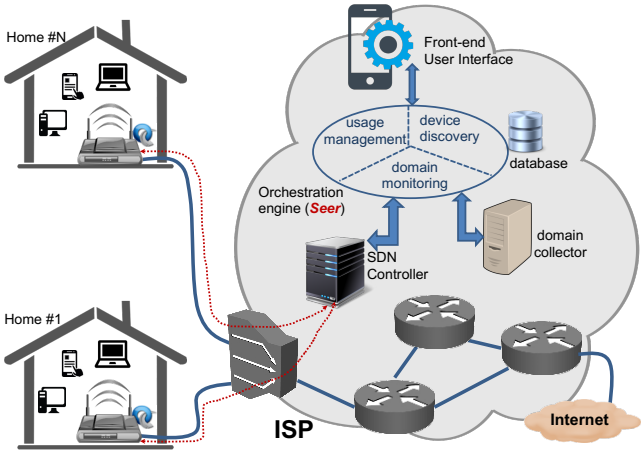



Fig. 1. System architecture

## II. Related Work

Prior works on analyzing home networks can be categorized as performance monitoring [1], [4]–[6] (e.g. broadband link speed, delay or loss rate); traffic characterization [2] (e.g. temporal pattern of HTTP, HTTPS or DNS applications), or troubleshooting [7] (e.g. remote diagnosis and repair). We aim at empowering users to better manage their home network (e.g. data consumption monitoring) and helping the ISP better understand household profile of online activity.

Various methods have been used to understand home networks characteristics. [4] measures broadband access link properties from outside the home, thus lacking visibility into what is happening inside home networks; BISmark [1] analyzes packets traversing the home gateway and characterizes the performance of home networks; [5], [8]–[11] run tools on end-host devices to measure broadband link properties. Many of these studies were based on one-shot measurements. We instead analyze home networks by continuously monitoring the home gateways.

The closest work to ours is BISmark [1] which embeds a custom firmware into the home gateway that performs both passive and active measurements including availability/capacity of the broadband link, usage of WiFi channels, number of devices, and packets statistics. BISmark gateways send home measurements to a cloud-based server every twelve hours. In our system, we do not customize the home gateway in any way whatsoever; this approach avoids vendor lock-in (we have tried multiple models of home gateways from multiple vendors), and leverages open-source firmware (OpenWRT and OVS) that is well-supported in the community. Further, we don't inspect and analyze packets. Instead, home gateways are dynamically configured by device-specific flow rules that provide visibility into devices' presence, their data usage volume, and domains they visit. Our SDN applications achieve this by pulling flow statistics from home gateways every minute.

## III. Home Network Visibility and System Architecture

We briefly describe visibility services that can be provided to benefit consumers and the ISP. We then outline our solution architecture.

### A. *Visibility for the Home*:

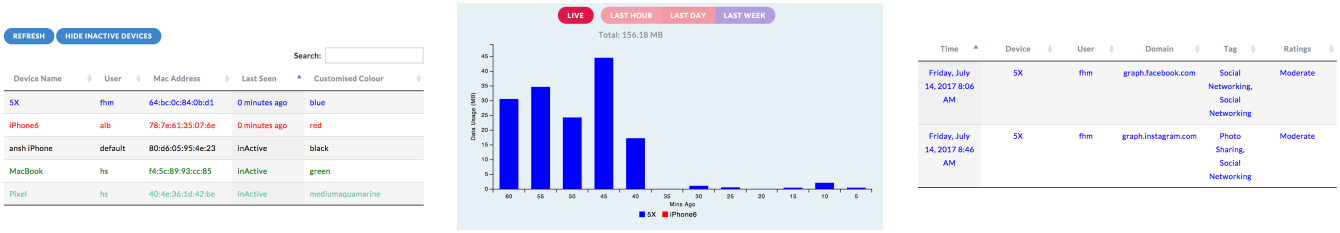
**Usage management and Device discovery:** Many users, mostly in developing countries, want to monitor data consumption of their individual devices due to a tight monthly quota imposed by the ISP as part of the Internet plan. This might help the user determine how to adapt usage pattern in the house. Further, consumers have little visibility into the number of connected devices, or the last time a device was seen in their home network. On the other hand, the ISP receives a significant number of support calls from its subscribers related to quota exhaustion – these subscribers often contest their usage, leading to complaints and disputes. The remedy for such situations can be as simple as giving the subscriber visibility into devices connected in their home, their presence on the network, and their individual data consumption on an hourly or daily basis. This way they can identify culpable devices, and if desired impose a limit on specific devices so that they do not exhaust the household's entire quota for the month.

**Domain monitoring:** Youth Internet addiction is becoming a serious concern in many developed countries, and is starting to afflict the developing world as well. In the US, kids aged 9-18 routinely spend several hours online daily, much of it unsupervised, and 70% admit to hiding their online activity from parents [12]. Though a plethora of client-side tools are available to shield kids from inappropriate content, including child-safe DNS resolvers, search engines, browser filters, operating system modes, and free/paid software suites, their uptake is poor as they demand high motivation from the parent to install, configure and maintain. The ISP has an opportunity to fill this gap, by providing the consumer with a single tool for monitoring of visited domains (i.e. content monitoring) across all household devices.

### B. *SDN Architecture*:

We now argue that SDN provides the best means to enhance above visibility features into home networks, both from a technology and business perspective. Our SDN-based architecture is depicted in Fig. 1. Data flow is shown by solid blue lines, and is conventional, i.e. from the home gateway to the DSLAM and through the ISP network out to the Internet. Control flow, which is what implements the new capabilities above, is depicted via dotted red lines. An SDN controller manages the home gateways, and applications providing the visibility features (i.e. usage management, device discovery, domain monitoring) reside on the controller and are exposed to users via a web-interface. The various components are described and justified below:

**Home router:** It is very tempting to embed the above features into the home gateway itself – as argued earlier,




(a) Device discovery. (b) Last hour usage. (c) Domain monitoring.

Fig. 2. User visibility into (a) connected devices, (b) hourly usage, and (c) domains visited.

we believe this is not the right approach. ISPs rarely design or manufacture home routers themselves, and instead partner with one or more vendor suppliers. Embedding custom-built features into the home router would lock the ISP with a supplier, increasing risk and cost if the supplier under-performs or over-charges. Further, embedded features are hard to upgrade once deployed, since that requires end-user consent. Lastly, configuring embedded features is cumbersome for end-users, who struggle with even basic instructions (such as logging in to 192.168.1.1) that are non-intuitive to the lay person.

We therefore strongly believe that the user-premises equipment should be an off-the-shelf device, that can be controlled externally using a standard well-understood interface. This is exactly what SDN provides. We have successfully taken commodity home gateways (incorporating a router and access point) and flashed them with OpenWRT (open-source firmware that works on a wide range of home gateways) and OpenVSwitch (open-source virtual switch along with OpenFlow agent), allowing them to function as generic switches whose forwarding behavior can be manipulated by an external controller. We emphasize that all the control logic now resides in the cloud, allowing continual updates and customization of features and user-interfaces without requiring any modifications to the home gateways.

**Controller:** The controller provides the substrate on which SDN applications reside and interact with the network devices (home gateways). They provide functionality such as maintaining the list of connected devices, inserting/deleting flow-table entries in switches using the OpenFlow protocol, and polling statistics. Several open-source controllers are available for this purpose, any of which can be used. We have chosen to work with FloodLight, but have written wrappers for its functions so that our application code is agnostic to the controller, allowing us to replace the controller at any point.

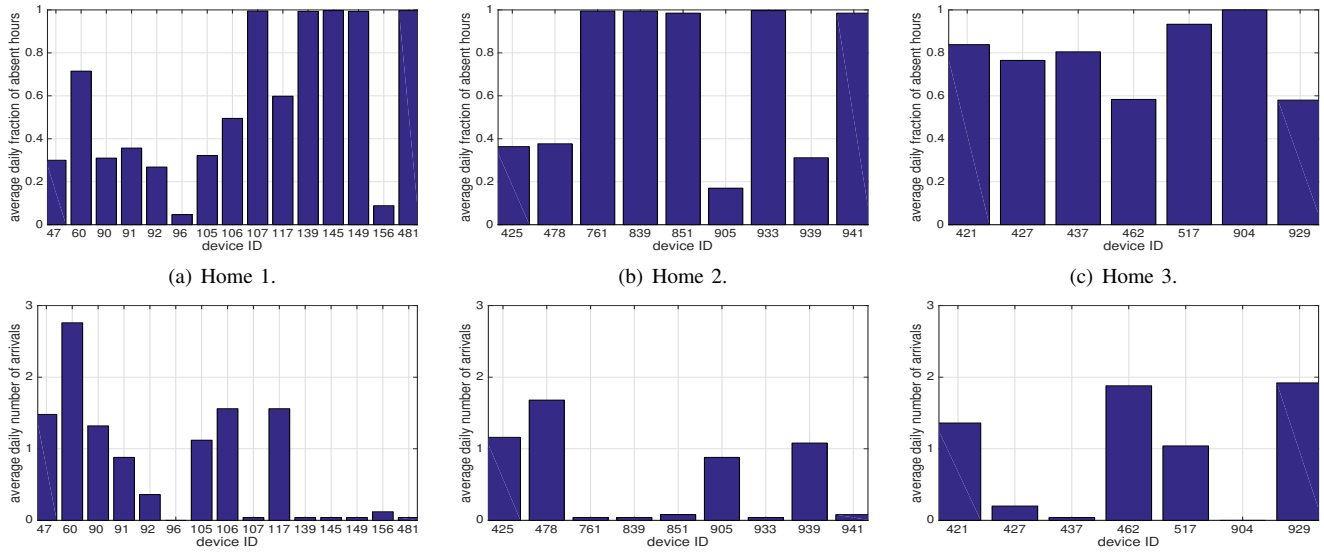
**Domain collector:** The domain collector receives a copy of all DNS queries generated by devices whose domain monitoring feature has been enabled by the user. DNS packets of user-enabled devices are mirrored to the collector using specific flow entries. The collector then obtains the tag corresponding to the domain name from OpenDNS, and posts the domain name, tag and the device MAC address along with its timestamp to the database.

**Orchestration engine and database:** The bulk of our intellectual property lies in the orchestration engine, which we call *Seer*. Seer implements the logic for the device discovery, usage management, and domain monitoring functions mentioned above, along with the databases of subscribers, their devices, preferences, usage, etc. It maintains the association between subscribers and their home gateway's data-path ID, so that control commands can be sent to the appropriate device. Its interactions with the controller (to the south) and the user-interface (to the north) are via REST interfaces, so that the implementations of the various components is decoupled (e.g. one SDN controller can be replaced for another, and the web-based user-interface replaced by a mobile-app). The frameworks, programming languages, and database schemas for the Seer are left entirely up to the ISP to choose, and are not constrained by the architecture in any way (as they would be if the features were embedded into the home gateways). The features themselves are implemented entirely in software operating in the control plane: for example usage management periodically polls the per-device counters from the home gateway to determine data usage, and domain monitoring receives a mirror of DNS queries to log visited domains and identify their tag (e.g social networking or video sharing).

**User interface:** The decoupling of features from the hardware platform permits arbitrary interfaces to be built: our current interface is web-based and interacts with the orchestration engine above using REST APIs; in the near future we intend to develop mobile apps that give a different experience for users but use the same underlying REST APIs. Yet again, we emphasize that the user-interface is in the cloud and not embedded into the home gateway (e.g. 192.168.1.1), so can be updated, improved, and customized at-will by the ISP.

Snapshots of our web-based user interface are shown in Fig. 2. In Fig. 2(a), we show a sample illustration of the web interface showing 5 ***devices*** for this trial user. The device name is auto-detected based on the DHCP hostname, but can be edited by the user via the interface. Each device can also be mapped to a user; an unmapped device is assigned to user "default". The table also shows the time the device was last seen on the home network. For example, in Fig. 2(a), it can be seen that there are currently two active devices on the network (those with last seen of 0 minutes ago) and other three devices are inactive.

The user interface has a tab that shows ***usage*** statistics, as depicted in Fig. 2(b). This tab can show instantaneous bandwidth usage ("Live") for each active device as well as




(a) Home 1. (b) Home 2. (c) Home 3.

(d) Home 1. (e) Home 2. (f) Home 3.

Fig. 3. Profile of daily presence

aggregate data volume (download plus upload) for each device over the past hour, day, or week, using a stacked bar-graph – we show the last hour usage in Fig. 2(b). This directly allows a subscriber to see which devices in their home are culpable for depletion of the monthly quota. Lastly, our user interface has a tab for ***domain monitor***, as shown in Fig. 2(c). This allows users to track domains (and their tag provided by OpenDNS) visited by devices. We note that this feature needs to be enabled by the user for devices of interest via the configuration tab of the web interface. Given a domain tag, we also provide rating to classify each visited domain with regard to suitability for audiences in terms of issues such as sex, violence or other types of mature content.

## IV. Analyzing Household Devices, Activity and Content

### A. Data collection

We have collected data from three homes (of this paper's authors) for a duration of 4 weeks, from May 1, 2017 to May 28, 2017. In order to maintain privacy of users, we call the three households "Home 1", "Home 2" and "Home 3". During our study, these homes have been hosts to a total of 31 devices. Our Seer engine scans individual home gateways every minute to collect statistics of device-specific flow entries that reveal the presence of devices, and their corresponding data consumption. DNS queries of user-enabled devices are captured by the domain collector.

### B. Data Analysis

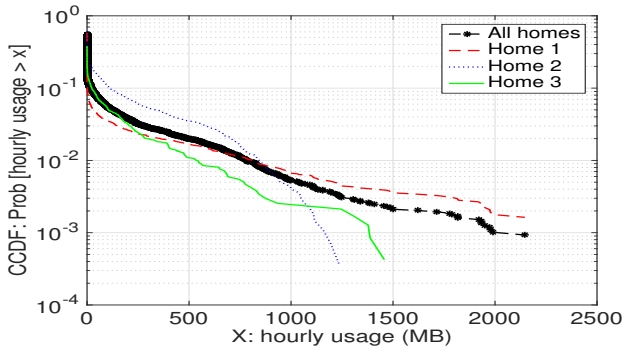
We now anlayze the data collected from these three households to get visibility into connected devices, their usage pattern, and domains they visited, with an hourly granularity (i.e. 24-hour clock convention).

*1) Connected Devices:* We aim to identify if a given device belongs to a person in the household (e.g. personal phone or laptop), or is shared among the family (i.e. TV or printer), or belongs to a visitor. The Seer tracks the activity of devices every minute. If a device is inactive for an entire hour (say, 9am-10am), we deem it to be "absent" for that hour. In other words, even one minute of activity for a device causes its status to be "present" in that hour. A change in device status from absent to present implies that the device has "arrived" to the network.

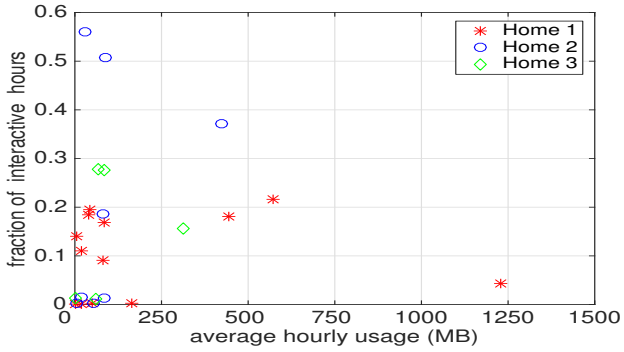
We plot in Fig. 3 the average daily fraction of absent hours and the average daily number of arrivals for each device in the three households. We expect devices which are always connected to the network, i.e. permanent and non-portable devices (generally shared by the family), to have a lower fraction of daily absent hours. For example, devices 96 and 156 in Fig. 3(a) are usually present and active in the home (i.e. they are rarely absent). These devices correspond to TV and printer in the Home-1 respectively. While considering Home-2 and Home-3 in Figures 3(b) and 3(c), we observe that household devices are absent for at least 4 hours a day (i.e. 18% of time). Note that devices with high fraction of absent hours (i.e. close to 1) correspond to visitor devices (for example, device 481 in Home-1, 941 in Home-2, and 904 in Home-3) – this was confirmed by individual households in our study.

Looking into arrival pattern, we expect personal devices of each household to arrive to the network on a daily basis. We see that six devices of Home-1, three devices of Home-2, and four devices of Home-3 arrive to their network at least once everyday, as shown in Fig. 3(d)-3(f). Unsurprisingly, permanent and visitor devices in all households have low average daily arrival values.

*2) Characteristics of Data Usage:* We then analyze data usage statistics of the devices to identify devices which consume high volume of data and the hours in which they do so. Devices that are connected to the network but not interacting much with a user (e.g. a printer) autonomously generate low volume traffic. In our analysis, we deem a device to be "interactive" over an hour, if it exchanges more than 1 MB data.




(a) CCDF of hourly usage




(b) Average hourly usage vs. interactive hours

Fig. 4. Profile of data usage: (a) CCDF of devices hourly usage, and (b) Average hourly usage versus interactive hours

TABLE I
NUMBER OF DOMAIN QUERIES.

| | total | background |
|---|---|---|
| Home-1 | 95,569 | 31,868 |
| Home-2 | 64,545 | 31,410 |
| Home-3 | 19,102 | 6,359 |

Fig. 4(a) shows the complementary cumulative distribution function (CCDF) of the hourly data usage of households devices for the entire duration of our study. We see that devices in Home-2 are more likely to have high volume hourly usage (i.e. more than 500MB) – probably watching videos. Though, only in Home-1 hourly usage of 1.5 GB is observed – implying that a large screen device (e.g. TV) plays videos .

Fig. 4(b) shows a scatter plot of usage versus interactive hours for households devices. It is seen that there are five data-hungry devices in the three homes, consuming a fairly large volume of hourly data on an average (i.e more than 250 MB). This implies that one device in each of Home-2 and Home-3, and three devices in Home-1, are used for watching videos frequently. Unlike other data hungry devices, the one in Home-2 (shown by blue circle) seems to interact with users more often.

*3) Profile of Visited Domains:* Finally, we aim to profile devices based on their visited domains. We note that visiting a domain causes a number of DNS queries to be generated from the device depending on its content. For example, visiting Facebook or Youtube generates tens of requests to dependent or advertisement domains. These subsequent queries (i.e."background" domains) are automatically generated while loading the page, but are not user initiated. We therefore filter them in our analysis. As shown in Table I, a significant fraction of DNS queries contribute to background domains, i.e. 33%, 48% and 33% in homes 1, 2, and 3 respectively.

## V. INSIGHTS AND INFERENCES

We now discuss insights that the ISP can gain from analyzing data collected from the subscriber's home network. We start by identifying the number of devices and their composition (i.e. shared, personal or visitor) in each household. Intuitively, the composition of devices can be identified by considering two metrics: the average daily fraction of absent hours, and the average daily number of arrivals. Since shared

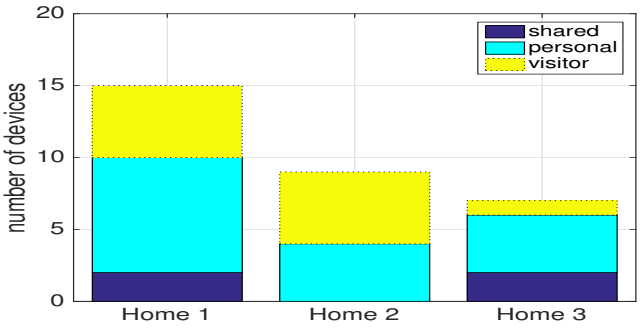
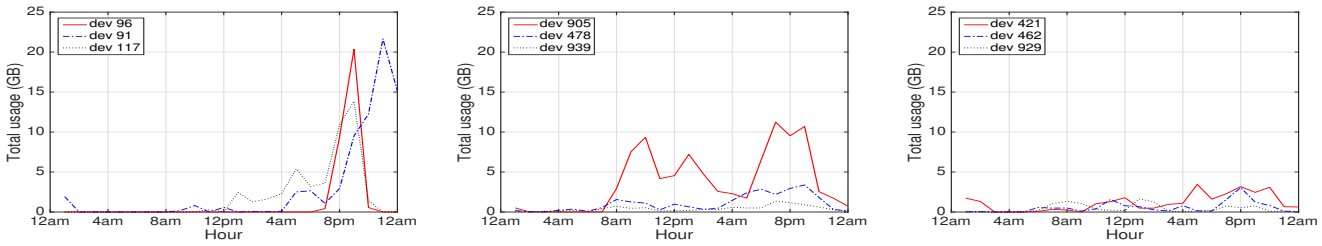



Fig. 5. Device composition in each household

(permanent) devices are present in the home network most of the time, we expect them to have low absent hours and low number of arrivals. Personal devices, on the other hand, are moderately absent (since they are carried outside home by their user on daily basis) and arrive to the network at least once a day on average. Lastly, visitor devices rarely become present on the network (i.e. high absent hours) thus their average arrival to the network becomes low too. We depict in Fig. 5 the composition of devices in three homes. It can be seen that Home-1 has more shared and personal devices compared to other households, implying a technology-hungry family.

We continue by inferring video viewing pattern in each household. We have chosen three data-hungry devices from each home network. In Fig. 6, we plot the total monthly volume of usage in each hour consumed by these data-hungry devices. We see a prominent pattern of high data usage, indicating video viewing, in Home-1 as shown in Fig. 6(a). Devices 96 and 117 are likely to consume videos between 9-10pm, whereas device 91 tends to play videos one hour later, i.e. between 10-11pm. Similarly, in Home-2 (Fig. 6(b)), device 905 seemingly consumes video contents at various hours of day whereas device 478 is used for watching videos during evening hours. By contrast, in Home-3 as shown in Fig. 6(c), a diurnal video watching pattern is not significant enough to be observed.

Finally, we select a personal device from each household and try to infer the content preference of their users. Fig. 7 depicts word clouds of visited domains (top row) and their corresponding tags (bottom row). We now identify those domains that are not commonly used. According to Figures 7(a)




(a) Home 1. (b) Home 2. (c) Home 3.

Fig. 6. Profile of data hungry devices


(a) Home 1, Device 117. (b) Home 2, Device 478. (c) Home 3, Device 462.

(d) Home 1, Device 117. (e) Home 2, Device 478. (f) Home 3, Device 462.

Fig. 7. Profile of contents by: (a-c) domain names, and (d-f) domain tags.

and 7(d), we observe that the device 117 in Home-1 visits special contents such as gaming (e.g. `minecraft.net`) and educational (e.g. `stanford.edu`) domains thus making it more likely to be owned by a teenage member of the family. Similarly, we infer from Figures 7(b) and 7(e) that device 478 in Home-2 is more likely to belong to an avid female social media user, since social media (e.g. `facebook.com`, `instagram.com`) and shopping (e.g. `taobao.com`) domains are visited frequently. Lastly, one can say device 462 in Home-3 belongs to a software-technology enthusiast referring to Figs. 7(c) and 7(f).

## VI. Conclusions

The explosion in Internet-connected consumer devices and the growing demand for Internet data is creating new pain-points for households grappling with an increasingly complex home network. We have argued that consumers and ISPs both can benefit from visibility into home networks. To-date, such capability has not been provided due to technological and economic reasons. We have presented an SDN-based architecture in which the features are built and operated in the cloud, and the home gateway is relegated to an off-the-shelf device running open-source firmware. We have deployed our system in three households and presented our analysis on how we can enhance visibility into device-level activity such as composition of household devices, their video viewing patterns, and content preferences. We believe that this provides some validation that both consumers and ISPs can benefit from increased visibility into the SDN-enabled homes, and hope that our approach gains wider acceptance in the near future.

## References

[1] S. Grover et al., "Peeking behind the nat: An empirical study of home networks," in *Proc. ACM IMC*, Barcelona, Spain, 2013.

[2] K. Xu, F. Wang, L. Gu, J. Gao, and Y. Jin, "Characterizing home network traffic: An inside view," *Springer Personal Ubiquitous Comput.*, vol. 18, no. 4, pp. 967–975, Apr. 2014.

[3] J. Yang and W. Edwards, "A Study on Network Management Tools of Householders," in *Proc. ACM SIGCOMM HomeNets*, New Delhi, India, Sep 2010.

[4] M. Dischinger, A. Haeberlen, K. P. Gummadi, and S. Saroiu, "Characterizing residential broadband networks," in *Proc. of ACM SIGCOMM Conference on Internet Measurement*, San Diego, CA, USA, 2007.

[5] M. Chetty et al., "Why is my internet slow?: Making network speeds visible," in *Proc. SIGCHI Conference on Human Factors in Computing Systems*, Vancouver, BC, Canada, 2011.

[6] R. Mortier et al., "Control and understanding: Owning your home network," in *Proc. COMSNETS*, Banglore, India, Jan 2012.

[7] E. S. Poole, W. K. Edwards, and L. Jarvis, "The home network as a socio-technical system: Understanding the challenges of remote home network problem diagnosis," *Computer Supported Cooperative Work (CSCW)*, vol. 18, no. 2, pp. 277–299, Jun 2009.

[8] L. DiCioccio, R. Teixeira, M. May, and C. Kreibich, "Probe and pray: Using upnp for home network measurements," in *Proc. of the 13th International Conf. on PAM*, Vienna, Austria, 2012.

[9] M. A. Sánchez, J. S. Otto, Z. S. Bischof, and F. E. Bustamante, "Trying broadband characterization at home," in *Proc. PAM*, Hong Kong, China, 2013.

[10] C. Kreibich, N. Weaver, B. Nechaev, and V. Paxson, "Netalyzr: Illuminating the edge network," in *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, Melbourne, Australia, 2010.

[11] L. DiCioccio, R. Teixeira, and C. Rosenberg, "Measuring home networks with homenet profiler," in *Proc. PAM*, Hong Kong, China, 2013.

[12] CNN. (2012) Survey: 70% of teens hide online behavior from parents. http://www.cnn.com/2012/06/25/tech/web/mcafee-teen-online-survey/.