Home Shield IoT Traffic Analyzer: A Comprehensive Analysis of Data Sharing Practices

Dhananjai Bajpai
Marquette University
dhananjai.bajpai@marquette.edu

Keyang Yu
Marquette University
keyang.yu@marquette.edu

***Abstract*—Internet of Things (IoT) devices have been expanding rapidly and significantly improved the automation and convenience in modern smart homes. Such functionalities are supported by large amount of data collection, analysis and sharing, which may bring privacy threat to the smart home users. It is crucial to identify unauthorized traffic volume data generated by IoT device, to help user better understand the privacy threat to their IoT environment. This paper presents a cost-effective approach to monitoring data-sharing activities of household IoT devices using the Cisco OpenDNS platform. We have analyzed the Internet traffic data generated from four popular devices to identify unauthorized third-party data sharing. We have discovered that such data sharing exists in multiple types of IoT devices installed in the smart home, the Smart TVs are sharing user-specific viewing data with third parties without user's consent, iPhone exhibits involuntary synchronization, and the IoT Plugs also show no unauthorized connection behavior. This user-specific, deployable pipeline contrasts with prior testbed-dependent studies and highlights the need for transparent data governance.**

I. Introduction

The Internet of Things, as one of the most important concepts in modern smart facilities, including smart homes, smart cities, smart factories, and even in the domain of healthcare, has revolutionized modern living by seamlessly integrating smart devices into our daily routines. Data collection, analysis and sharing from IoT devices have increased their level of autonomy. However, this connectivity comes with the price of privacy and security risks. Behind their user-friendly interfaces, IoT devices often engage in opaque data-sharing practices, leaving users vulnerable to unauthorized data collection and transmission.

Unfortunately, many IoT device manufacturers choose not to disclose their data-sharing pipelines, leaving users with limited visibility into whether their personal or household data is being transmitted to third-party servers, under what circumstances, and to whom. Such uncontrolled data sharing may lead to data ethical problems and data privacy concerns.

Workshop on Innovation in Metadata Privacy:
Analysis and Construction Techniques (IMPACT) 2025
28 February 2025, San Diego, CA, USA
ISBN 979-8-9919276-2-8
https://dx.doi.org/10.14722/impact.2025.23051
www.ndss-symposium.org

To address this problem, we introduce a household-specific approach to IoT traffic analysis, addressing the gap left by prior research, which often relies on generalized or public datasets. These datasets cannot capture the unique behavior of individual devices within specific contexts, as device settings and usage vary significantly across households. By empowering users to collect and analyze their own IoT traffic data, this work offers a personalized methodology to identify suspicious or unethical data-sharing practices, evaluate device trustworthiness, and highlight the importance of transparency in IoT ecosystems. Major motivation for this project is driven by the need to uncover the "behind-the-scenes" activities of smart devices for ensuring data privacy.

The following sections will be structured as follows:

Background: We evaluated existing approaches on their data collection pipelines and listed their limitations on abnormal traffic capturing. This motivates our work on designing a user-friendly, secure and flexible IoT network traffic analyzer for better preventing user privacy leakage from uncontrolled connections.

Design: We present the framework for capturing data from popular smart home devices including the LG Smart TV, Apple iPhone, IoT Plug, and TP-Link Router in four operational states: power-on routine, idle state, regular usage, and shut-down routine.

Evaluation: By analyzing the traffic patterns, the study identifies both voluntary user actions and involuntary background activities, exposing ethical concerns such as unauthorized data sharing.

II. Background

In this section, we summarize prior approaches to monitoring abnormal smart home network traffic. Sivaraman et al. examined the security and privacy implications of smart IoT devices in homes, revealing numerous vulnerabilities in the way these devices handle user data [5]. Similarly, Hamza et al. proposed using Manufacturer Usage Description (MUD) profiles to verify and monitor the behavior of IoT networks, in order to improve transparency and security [1]. Pashamokhtari et al. introduced efficient IoT traffic inference techniques using multiview classification and progressive monitoring, which

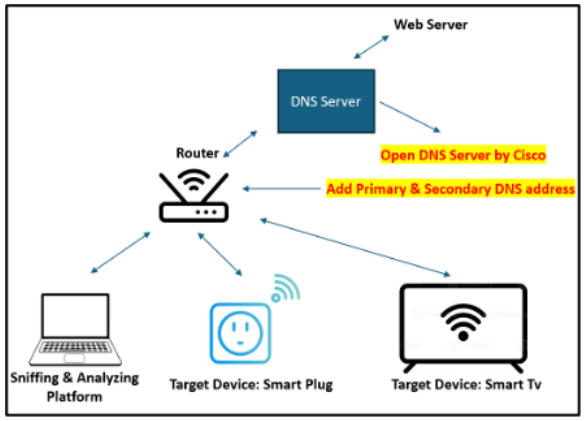

Fig. 1. The architecture of data collection pipeline.

Fig. 2. Cisco's umbrella setup.

allow for more adaptive identification of data flows [2]. Furthermore, Pashamokhtari et al. explored a combined stochastic and deterministic modeling approach to infer connected IoT devices, which, while effective, still faces challenges in adapting to dynamic and unpredictable behavior in real residential environments [3]. Clustering-based network traffic modeling, inspired by the work of Sivanathan et al., was adapted to classify device behavior, allowing the identification of unusual activity that could indicate unauthorized data sharing or other threats [4]. The key limitation of these existing approaches is that their data collection pipelines cannot be replicated by different users. These studies consume data generated from a complex test bed that relies on various hardware devices. Contrary to previous studies, this paper proposes an easy-to-replicate data collection pipeline at no cost which can be implemented by any user within minutes.

III. Design

In this section, we first list the major design challenges for the proposed IoT traffic collector, followed by a comprehensive overview of the setup of the data collection pipeline, targeted devices, the description of the collected data, data processing, and feature collection.

A. Design Challenges

The first challenge was limiting the cost of setting up the data collection pipeline. Using Cisco's open DNS umbrella, the user can avoid configuring complicated local data collection hardware, including smart routers and hub devices. The requirement for traditional traffic-capturing scripts like TCPDump or Wireshark, though proved to be effective, still requires additional efforts for regular smart home users.

The second challenge was distinguishing between normal and potentially malicious activities. Instead of using publicly available. predefined datasets, our data collection lacks precise labeling. By adding columns of user activity as groundtruth, the data could be identified as user-interactive data and non-user-interactive data as a substitution. User activities were denoted for the entire duration of the data collection process with the corresponding time stamps.

B. Data Collection Pipeline

We leveraged Cisco OpenDNS to reroute household Internet traffic for logging and analysis. All outgoing pings are captured and exported as a CSV file. Figure 1 illustrates the data collection architecture, showing the traffic flow and intercept points. The target devices, as well as a personal computer acting as data sniffing and analyzing platform are connected to the central router; all of the network traffic generated under this router will be first rerouted to the Cisco DNS server before the destination. This was achieved by configuring primary and secondary DNS addresses on the router.

Step-1: Account registration is required for the Cisco OpenDNS umbrella, as shown in Figure 2. Registration requires an organizational email and provides a 14-day trial, after which a subscription is needed to maintain access. This step ensures that all relevant network activity is captured for analysis.

Step-2: The primary and secondary DNS keys must be configured in the router, as shown in Figure 3. These keys direct Internet traffic through OpenDNS, which allows for monitoring. Depending on the router, these keys are entered into the WAN or DHCP settings, ensuring that all traffic flows through OpenDNS for visibility into device communication.

Step-3: Then we will need to configure the DHCP settings with the DNS keys on the router, as shown in Figure 4. After the configuration, the router will transfer Internet traffic to OpenDNS, enabling real-time traffic monitoring. The OpenDNS dashboard allows data visualization and downloading directly.

This streamlined process allows detailed monitoring of network activity with minimal effort, ensuring both voluntary and background device communications are captured for analysis.

Fig. 3. Primary and Secondary key configuration.

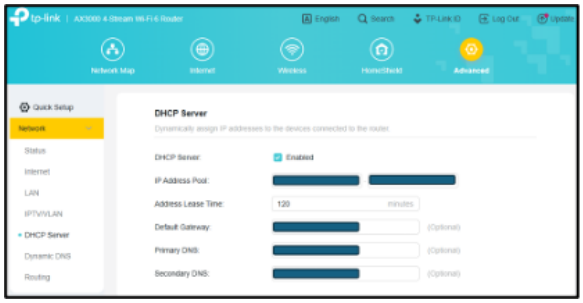

Fig. 4. Add Primary and secondary key in DHCP.

Date	Time	Destination	Categories	Application
10/22/2024	19:35:46	gateway.fe2.apple-dns.net	Infrastructure and Content Delivery	Apple iCloud
10/25/2024	21:24:39	disney.images.edge.bamgrid.com	Infrastructure and Content Delivery	Bamtech Media
10/25/2024	21:24:49	home.hulu.com	Movies, Television, Streaming Video	Hulu
10/25/2024	21:24:43	logs.netflix.com	Movies, Television, Streaming Video	Netflix

Fig. 5. Features collected via data pipeline.

C. Targeted Devices

In this paper a total of four smart home devices are chosen for experiments. All devices were connected to, used for 14 days and removed from the TP-Link router.

D. Data Description

Cisco Umbrella captures and stores the Internet traffic data flowing through a targeted router in a csv format. Figure 5 shows the unique features captured and saved in corresponding columns respectively. The time span for data collection is 14 days. The targeted devices have four major status including switched on, switched off, left in ideal state, and periodically browsed.

E. Data Processing

The captured data are pre-processed to ensure integrity by removing rows with missing values or corrupted strings. A ”User Activity” column is added to the dataset to link IoT device actions with user inputs. For example, when watching Netflix, the column is marked as ”Using,” while it is labeled ”Idle” when the Smart TV is inactive. Similarly, device states like ”Power-On” and ”Power-Off” denote startup and shutdown events. For the IoT Plug, activities are labeled as “IoT Plug On” and “IoT Plug Off” to conceal behaviors of each device, including power-on routines, power-off routines, idle states, and general browsing patterns. its behavior. This categorization enables clear identification and correlation of IoT activities across devices.

F. Feature Selection

Key features include ”Date,” ”Time,” ”Destination,” ”Categories,” and ”Application.” The ”Date” and ”Time” fields are merged and re-sampled at 12-hour intervals to identify websites and applications accessed by devices over time. The ”User Activity” column distinguishes user-authorized actions from background activities. The ”Application” column highlights frequently visited services, while the ”Destination” column ranks accessed HTTPS links by frequency, providing insights into device behavior.

IV. Results

This section presents the results of the analysis of network traffic data for the four targeted devices, using appropriate visual representations, such as graphs and charts. The goal is to understand both the visible and hidden behaviors of each device, including power-on routines, power-off routines, idle states, and general browsing patterns.

A. Smart TV

The smart TV exhibits distinct behaviors during power-on, depending on the last app used before shutdown. Initiate background processes to resume or update applications, performing additional activities related to app management and synchronization. Two power-on scenarios were observed shown in Figure 6.

Case-1: When powered off from the home screen, the TV reloaded it upon power-on, sending pings to Alphonso for ad optimization.

Case-2: When powered off while using Netflix, the TV updated Netflix logs, configuration files, LG system checks, and Alphonso during power-on. In both cases, the TV was automatically connected to an LG server based in the United States, proving the built-in synchronization routines.

1) Smart TV Malicious Data Sharing: The Smart TV frequently shared user data with Alphonso, a third-party entity, without user consent.

Figure 7 below shows Alphonso as the second most contacted destination over 14 days.

Data shared include video / audio content, IP addresses, location, viewing history, device details, and demographics,

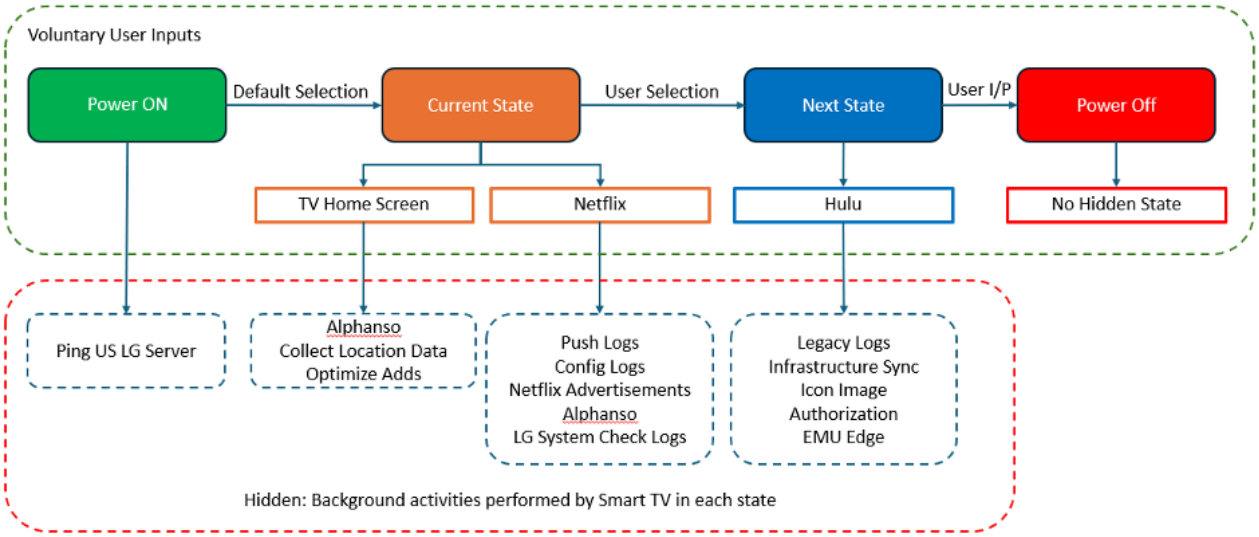

Fig. 6. LG smart TV power-on and browsing voluntary and involuntary routines.

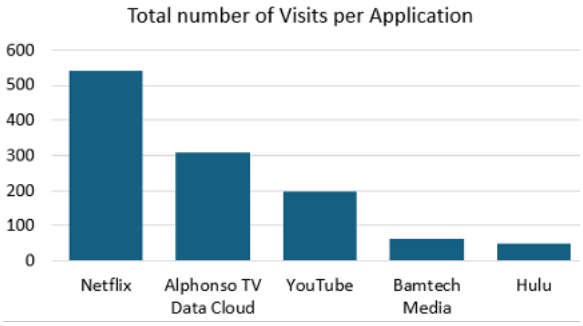

Fig. 7. Top 10 destinations visited by Smart-TV.

aggregated as "Viewing Data" and distributed globally as shown in Figure 8. This unauthorized data sharing violates privacy principles and lacks transparency, raising ethical concerns.

2) User TV Watching Habits: Figure 7 in the previous section also reveals that voluntary usage primarily involved Netflix and YouTube, Disney, and Hulu. The time spent on Netflix was twice that of any other TV application.

B. IoT Smart Plugs

IoT smart plug from "KASA" did not perform any out of ordinary or malicious activity. It performed only voluntary pings; no back-end or involuntary pings were found during data analysis. The number of voluntary clicks can be directly mapped to actual server pings, as shown in Figure 9.

C. iPhone

To isolate the behavior of the iPhone, all other devices were disconnected from the router. Figure 10 shows that Apple iCloud was accessed involuntarily at more than twice the rate of any site visited voluntarily, maintaining the consistency of the data for iCloud storage.

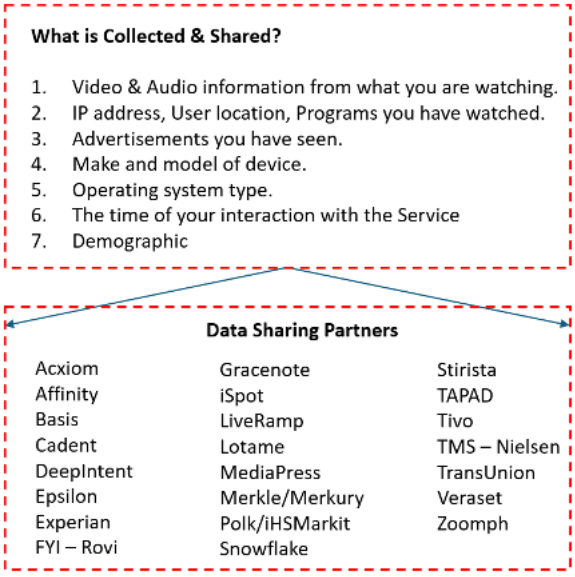

Fig. 8. List of variables collected from each user and list of data sharing partners.

Services such as iTunes, Maps, and Safari, though unused, generated continuous background pings. No suspicious or unauthorized third-party data sharing was identified, indicating involuntary synchronization occurred only within Apple's ecosystem.

D. TP-Link Router

The traffic of the TP-Link router, shown in Figure-11, revealed that Cisco's OpenDNS logged frequent pings to domains such as "EEROUP" and "E2RO" to maintain the integrity of the mesh network. "EERO" managed Wi-Fi connectivity through diagnostics and updates, while "node.e2ro.com"

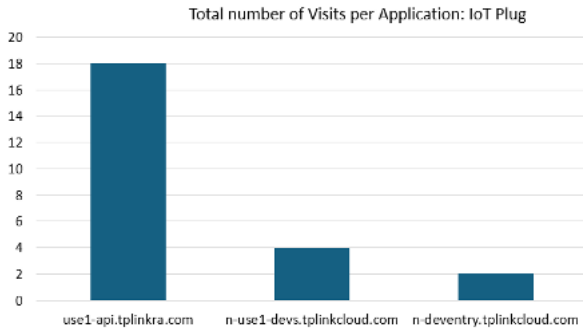

Fig. 9. Apps visited and total count by IoT Plug.

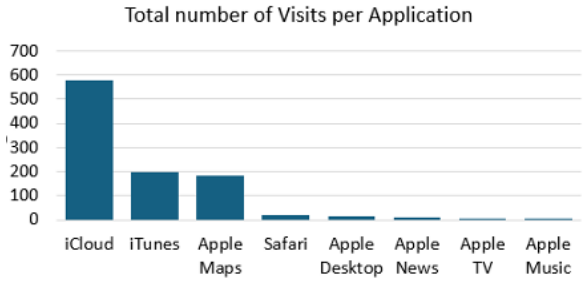

Fig. 10. Total count of involuntary Apple apps visited by iPhone.

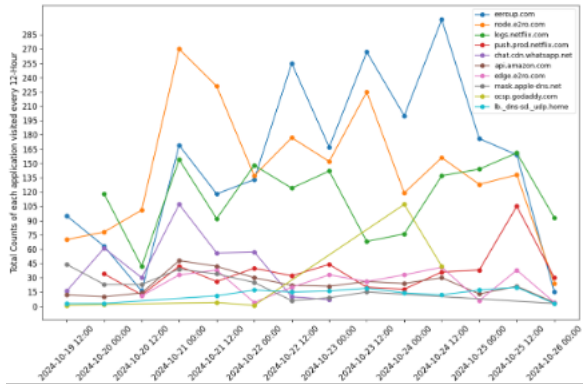

Fig. 11. Total traffic through router per 12 hours.

handled telemetry and health checks. Voluntary application traffic, led by Netflix streaming, was half as frequent as mesh Wi-Fi maintenance, reflecting the high overhead required for robust connectivity.

V. Conclusion and Future Work

This paper presents a cost-effective method for capturing household IoT traffic using OpenDNS, requiring no additional hardware. The analysis of four devices LG Smart TV, iPhone, IoT Plug, and TP-Link Router revealed that Smart TV shared user data and metadata with third parties without consent, violating privacy norms. However, other devices were not associated with such activities.

A key limitation of this study is the lack of visibility into detailed network metadata, such as packet sizes, source and destination ports, and transport protocols, which restricts insights into the volume and nature of data transferred, preventing a deeper understanding of involuntary data sharing practices.

Future work aims to extend the pipeline to analyze data-sharing practices of other Smart TVs, such as Samsung and Sony, enabling broader comparisons between IoT manufacturers. In addition, unsupervised machine learning models will be explored to detect anomalies and categorize device behaviors, enhancing the ability to identify unauthorized background data-sharing activities.

References

[1] Ayyoob Hamza, Dinesha Ranathunga, Hassan Habibi Gharakheili, Theophilus A. Benson, Matthew Roughan, and Vijay Sivaraman. Verifying and monitoring iots network behavior using mud profiles. *IEEE Transactions on Dependable and Secure Computing*, 19(1):1–18, 2022.

[2] Arman Pashamokhtari, Gustavo Batista, and Hassan Habibi Gharakheili. Efficient iot traffic inference: From multi-view classification to progressive monitoring. *ACM Trans. Internet Things*, 5(1), December 2023.

[3] Arman Pashamokhtari, Norihiro Okui, Yutaka Miyake, Masataka Nakahara, and Hassan Habibi Gharakheili. Combining stochastic and deterministic modeling of ipfix records to infer connected iot devices in residential isp networks. *IEEE Internet of Things Journal*, 10(6):5128–5145, 2023.

[4] Arunan Sivanathan, Hassan Habibi Gharakheili, and Vijay Sivaraman. Detecting behavioral change of iot devices using clustering-based network traffic modeling. *IEEE Internet of Things Journal*, 7(8):7295–7309, 2020.

[5] Vijay Sivaraman, Hassan Habibi Gharakheili, Clinton Fernandes, Narelle Clark, and Tanya Karliychuk. Smart iot devices in the home: Security and privacy implications. *IEEE Technology and Society Magazine*, 37(2):71–79, 2018.