Talk to Me: Investigating the Traffic Characteristics of Amazon Echo Dot and Google Home

Frank Loh, Stefan Geißler, Fabian Schaible, Tobias Hoßfeld
University of Würzburg, Institute of Computer Science, Würzburg, Germany
{firstname.lastname}@uni-wuerzburg.de

***Abstract*—The application area of smart speakers is steadily increasing with the Amazon Echo family and Google Home being brand leaders. Use cases include, among others, updating the calendar, home automation, or simply assisting users in their every day life. With the increasing amount of devices, the traffic generation and requesting process becomes a relevant subject to be studied in order to create traffic models and predict the impact on networks by these types of devices. Furthermore, with a detailed understanding of the devices, service quality can be monitored and improved for the end user. For that reason, in this work the requesting and traffic generation process of the Amazon Echo Dot and Google Home are studied and the generated network load of both devices is compared. With the insights of this study and additional device usage statistics, detailed traffic and usage models can be created.**

***Index Terms*—Home Automation, Smart Assistants, Traffic Analysis, Measurement**

I. Introduction

Internet of Things (IoT) is one of the major new trends in communication networks. Sensors, smart infrastructure, smart homes, and smart cities are only a few examples of application areas. In contrast to application layer characteristics, the network behavior of many of these devices has not been well researched in the past. However, understanding the details regarding their traffic characteristics is crucial for the creation of reliable traffic models, future network dimensioning, traffic forecasts, and service quality monitoring.

In recent years a new group of devices, so called smart speakers, became hugely popular. These devices can be used in a users every day life for, among others, updating the calendar, setting alarms, accessing the latest news, receiving up-to-the-minute weather forecasts, or even automatically processing a phone conversation. In short, understanding and processing a request spoken in natural language.

Taking a more detailed look at the responsibilities of smart speakers, the scope of challenges becomes visible. The constantly growing amount of features and apps, called skills, added to their application area, a totally heterogeneous amount of users, or the goal to satisfy all users regarding service quality and service processing times are just a few examples.

Since the introduction of the first smart speaker, the Amazon Echo, in June 2015 [1], sales figures have steadily risen. Within the first year alone, sales have more than tripled [2]. Brand leaders are by far Google and Amazon with a combined market share of 88.8 % on the US market [3] and 63.9 % globally [4].

From a communication researchers point of view, detailed knowledge about the behavior of devices is important for two reasons: First, insights into the traffic generation process of every device in a network is a valuable information for traffic model generation and traffic engineering. Second, analyzing the devices and their services in detail allows the detection of bottlenecks and delays influencing service quality.

In this work, we present traffic measurements conducted with the Amazon Echo Dot and the Google Home smart speakers. This is a first step towards classifying the expected load on networks and to better understand the traffic characteristics of these devices. During the study, different requests are sent to and processed by the devices while the resulting traffic is measured at network layer. We study the requesting behavior and delays of both devices and compare the results, both in uplink and downlink direction for different scenarios.

The contribution of this work is threefold: First, the behavior while requesting a specific task is studied. By comparing the generated traffic of the Amazon Echo Dot and the Google Home, we show that the Google Home creates on average more than two times more uplink traffic than the Amazon Echo Dot for a single request. Second, differences in request processing are detected and presented. Last, by detecting the processing duration of both devices, delays are examined and compared as valuable information for the service quality. This is, to the best of our knowledge, the first request and network traffic study of these smart speaker devices. The insights provided in this paper can be used as a basis to create more complex and detailed traffic models.

The remainder of this work is as follows: in Section II, background is presented and related work is summarized. Afterwards, in Section III, the testbed, the scenario selection, and the study is described followed by the discussion of the measurement results in Section IV. Section V concludes.

II. Background and Related Work

The following section provides the fundamental background required to understand this work. First, general information about smart speakers is summarized followed by a detailed focus on the Amazon Echo Dot and the Google Home device. At the end of this section, related work is discussed.

A. Background

The term *smart speaker* denoted a wireless audio playback device with the ability to connect to different types of audio

sources like media libraries on various platforms and online services. Typically, these connections are established via Wi-Fi or Bluetooth. Special focus lies on the preservation of the ease of use despite the numerous technically diverse connectivity options the devices offer so that everyone, even technically inexperienced users, can easily operate them. This ease of use has in recent years been improved and enhanced by the addition of voice-controlled intelligent personal assistants (IPAs), often also called virtual, digital, or AI assistants. Today, they might be the most integral and generally best known part of smart speakers. Therefore, this paper focuses on exactly these voice-controlled devices.

IPAs are applications that are able to understand natural human language which makes it possible to operate them via voice commands. Their purpose is to assist the user by facilitating and accomplishing various tasks. Common examples are the management and scheduling of personal appointments, giving weather forecasts, providing information users would normally look up on the Internet, as well as controlling connected compatible smart home devices. In order to be always able to respond to commands, most IPAs are always listening for their wake words unless being disabled. IPAs are typically composed of two parts. One is a piece of software which is installed on the user's device and serves as the user interface. The other one is the software which actually interprets the input and processes the tasks. The second part is provided online in a cloud. Thus, IPAs require an Internet connection in order to work. One major advantage thereof is that the complex and processing intensive tasks of voice decoding and interpreting do not have to be done by the user's device but are forwarded to specialized data centers with much higher processing power [5]. Another advantage is that the uploaded data can be used to improve different components like machine learning, natural language processing, and speech recognition. This helps to enhance the whole artificial intelligence behind the assistants, which in turn improves user experience and satisfaction. However, there is also a downside to it. The collection and storage of uploaded data constitutes a threat to the users' privacy as they have virtually no control over how their data is stored and processed [5].

B. Related Work

In this section, related work for traffic characteristic studies is summarized and compared to this work with focus on IoT technology and smart homes. Then, smart speaker specific works are presented with a main focus on the usage behavior.

To create a widespread traffic model in a smart home, all devices in this context must be monitored, and network characteristics analyzed. Thus, [6] presents insights in measurements of 28 different IoT devices. They show traffic characteristics of the different types of devices and discuss trade-offs between cost, speed, and performance. One step further is done in [7]. There, traffic characteristics of IoT devices in a Smart City context are mapped to energy consumption. A more general approach is done in [8], where IoT traffic is characterized in smart cities and campuses, while in [9], smart home IoT traffic is monitored passively. Thus, the authors see characteristics in the every day usage. Nevertheless, to the best of our knowledge, no work is available addressing the traffic generation and requesting behavior of smart speakers in detail at packet and request level. For that reason, we monitor generated network traffic, investigate the requesting behavior of the devices, and compare different content requests of the Echo Dot and the Google Home as the main representatives.

Studying smart speakers is often about the ability to understand and fulfill a wide range of spoken tasks, that has a main focus on user satisfaction. The authors of [10] for example are focusing on user satisfaction for the Amazon Echo devices. Often, frequent requests are studied, while infrequent users of such personal assistants receive special attention in [11]. With a deeper look at smart speakers, installed on a large amount of different types of devices, research exists that investigates and compares the general capabilities of smart assistants [12]. Bringing together the usage behavior of frequent and infrequent user, and the general capability of smart speakers, [13], [14] gives an overview about usage patterns valuable for model creation. Based on these usage patterns and the requesting behavior and size presented in this work, a detailed model about the traffic generation process of smart speakers can be created. Compared to this, the amount of downlink traffic and the traffic patterns to the devices is determined by the used application. This is widely studied by many works, like e.g. [15] studying the downlink behavior when streaming video with the Amazon Echo Show. Other works, like [16], are focusing on music streaming. Since this is not directly related to smart speaker features, and highly related to the application, it is not discussed in this work.

III. Measurement Setup and Scenarios

In this section, first the measurement methodology is described followed by the investigated scenarios and data post-processing steps.

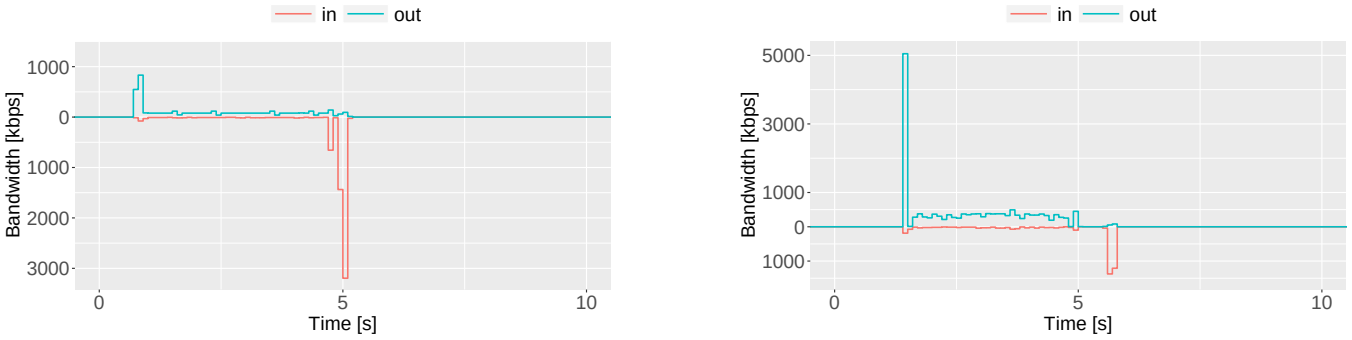
A. Testbed

For data capturing with the Amazon Echo Dot and the Google Home devices, an automated measurement setup is created. It contains both smart speaker devices, connected to a TP-Link TL-WR1043ND V16 Gigabit WLAN router via WiFi. On the router, the free Linux operating system OpenWrt is running in version 12.09-rc1. OpenWrt is optimized for embedded devices and provides a fully writable file system with package management. This enables the installation of the capturing software *tcpdump*. Additionally, a regular speaker is installed and connected to a computer to play pre-recorded voice requests. To avoid misunderstanding the spoken voice commands, the setup is created in an isolated environment at the University of Würzburg to minimize background noise and interference with the measurement.

For data collection, the testbed is established as follows. The router, that is connected to the Internet opens a WiFi access point for both smart speakers. For each measurement, a pre-recorded audio file is played to request different content

TABLE I
SCENARIO OVERVIEW

Scenario	Description	Wording
Baseline	Traffic investigation in the idle state	None
Wrong wake word	Addressing devices with wrong wake word	None
Music playback	Music playback with and without addressing the device	None
News request	Requesting daily news with addressing the device	None
Addressing only	The devices are only addressed correctly	Only wake word
Factual knowledge	The devices are addressed and asked a factual question	*Wake word, who is Alan Turing?*
Weather request	The devices are addressed and asked for the weather report	*Wake word, how will the weather be tomorrow in Würzburg?*



(a) Amazon Alexa

(b) Google Home

Fig. 1. Exemplary transaction of different home assistant devices.

from Amazon Echo Dot and Google Home. These files are automatically played from the computer via a connected speaker. The traffic generated between the devices and their corresponding cloud service endpoints passes the router, where it is captured by the tcpdump traffic capturing tool. For the evaluation, the source and destination IP-addresses, ports and packet sizes are extracted for flow separation and flow size determination. The packet arrival timestamp is logged for time based evaluations. In the following, all measured scenarios are outlined and summarized in Table I, before a detailed evaluation of a representative scenario is presented in Section IV. The scenarios are selected to cover the most used application areas according to [13].

B. Scenario Description

In the following, the different measurement scenarios conducted in the context of this work are described and a selection for more detailed evaluations in Section IV is made.

Initially, the traffic generated by the devices in the idle state is conducted in a *baseline* measurement. There, no interaction with the devices occurs while all traffic is captured. The results show a regular peak of, on average, 16 kB every roughly 24 h for the Google Home device and 26 kB every 8 h for the Amazon Echo Dot. Additionally, no significant amount of traffic has been measured. Thus, we will provide no additional investigation of this scenario.

Next, scenarios are measured to study the behavior of the devices if not addressed correctly or directly. In the *wrong wake word addressing* scenario, the devices are addressed with *Peter* instead of *Alexa* or *Okay, Google* respectively. Similarly, in the *music playback* scenario, music is played in the background without addressing the devices directly. The results show, that no additional traffic is produced if the devices are not requested directly or requested with the wrong wake word. Thus, we conclude that the devices do not process any background noise.

To study the traffic generated by addressing the device, the *addressing only* scenario is evaluated. We have observed that the process of leaving idle state when the devices are triggered with the wake word and going back to idle state if no actual request is sent follows a deterministic pattern. The amount of transmitted data per request amounts to 17 kB for Amazon Alexa and 33 kB for Google Home on average. Thus, we suggest an empty voice request is sent until the devices recognize that no content is requested. Furthermore, we detect the same behavior if requesting the devices only slightly wrong, for example with *Alex* in case of Alexa.

In order to study the addressing, requesting, and answering process, the *factual knowledge* scenario is measured. Here, the devices are asked *"Who is Alan Turing?"* to investigate the traffic variation for the same request, while the *wake word* is *Alexa* or *Okay, Google* respectively. The results show that similar requests lead to similar traffic patterns in the uplink and downlink between multiple repetitions. As the results of this scenario and the news request scenario are very similar to the *weather request* scenario, we omit the results of this

measurement for brevity reasons. Additionally, larger content requests, like investigated in the music playback scenario shows similar behavior in the uplink. Furthermore, since the downlink depends on the requested content, we refer to related work discussing content specific requesting e.g. for audio [16].

Finally, the *weather request* scenario is studied to evaluate the behavior of the devices when requesting the same information in different ways. To this end, the devices are asked *"How will the weather be tomorrow in Würzburg"*. To investigate the traffic variation in the uplink for requests of different duration, the length of the question is varied while the amount of information is kept constant. This is done with longer pauses in the requests, repetitions of single words, and additional filler words without changing the meaning or informational content. Request duration in this scenario varies between 3 s for the short duration request, 5 s for the medium duration and 10 s for the long duration request. For all request types, more than 70 repetitions are made to increase sample size and obtain statistically significant results.

IV. Evaluation

In the following section, the different characteristics of the transactions between a user and the home assistant systems covered in this work are investigated. For that reason, the weather request variations are analyzed in detail. To this end, we first identify different properties of a typical transaction as seen by the network.

A. General Request Structure

To identify the general properties of a home assistant transaction, we focus on a single exchange as seen on network level for both devices evaluated in this work. Figures 1a and 1b shows an exemplary transaction by Amazon Alexa and Google Home respectively. Thereby, the outbound bandwidth in positive Y direction in blue and inbound bandwidth in negative Y direction in red is presented. The x-axis shows the time in seconds.

It can be seen that both transactions follow the same pattern. The transactions start with an initial burst of outbound data, followed by a continuous stream of outgoing traffic. This pattern is explained by the behavior of the smart devices, as the devices start to buffer voice data while establishing a connection to the cloud based service endpoint. The initial burst represents the pre-buffered data. The following traffic represents a continuous stream of voice data while the user speaks to the device. Analogously, the transaction ends with a short burst of incoming data, that resembles the reply streamed from the cloud service.

Based on this initial observation, we now evaluate the different characteristics of this transaction pattern.

B. Transmission Size

To study the transmission size, Figures 2a-c show the cumulated data in downlink and uplink direction as boxplots for three different transaction scenarios, with more than 70 repetitions each. Namely the request for a weather report as described in Section III.

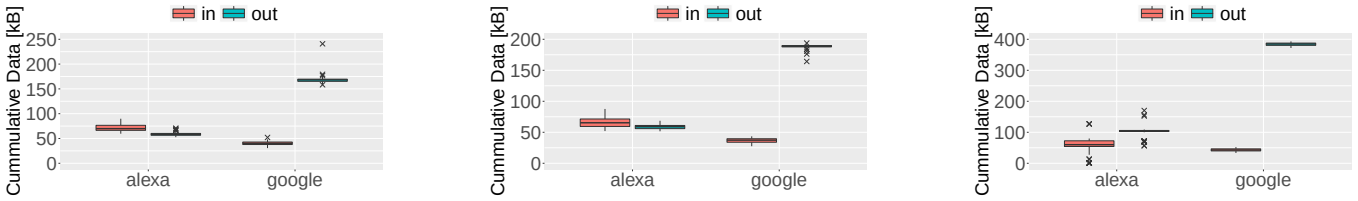
Inbound data is depicted in red, outbound data in blue. The black markers mark outliers whose distance to the mean is either smaller than $Q1-1.5\cdot IQR$ or larger than $Q3+1.5\cdot IQR$, where IQR is the inter quartile range, $Q1$ is the 25% quantile and $Q3$ is the 75% quantile.

It can be seen that the total values are in the range of a few hundred kilobytes and the variance within the different scenarios is low. This is expected since the same request in general leads to the same reply. Nevertheless, we see some outliers, especially for the long duration request phrase in case of Amazon Alexa in both inbound and outbound direction. The inbound lower end outliers stem from repetitions in which the device was not able to establish a connection to the service. This can be a result from extending the request duration without adding additional content. The upper end outliers suggests that significantly more data was downloaded in this repetition. Most likely, the system suggested further functionality that the assistant system can perform for the use, which led to additional speech data that needed to be downloaded. The outbound outliers in both directions are explained by the behavior of the device when a request is not comprehended correctly. The device sometimes shuts off ahead of time or listens on for a few moments, although the request was already spoken completely. Finally, it can be seen that the Google Home device generates three to four times more outbound data when compared to Amazon Alexa. This is most likely due to a higher quality audio codec, since Amazon Alexa uses 48 kbps constant bitrate encoding while Google Home can use up to 96 kbps [17], [18]. The inbound traffic is similar, independent on the requesting duration, while a little lower for the Google Home. This is obvious, since the same content is requested.

C. Transaction Duration and Processing Time

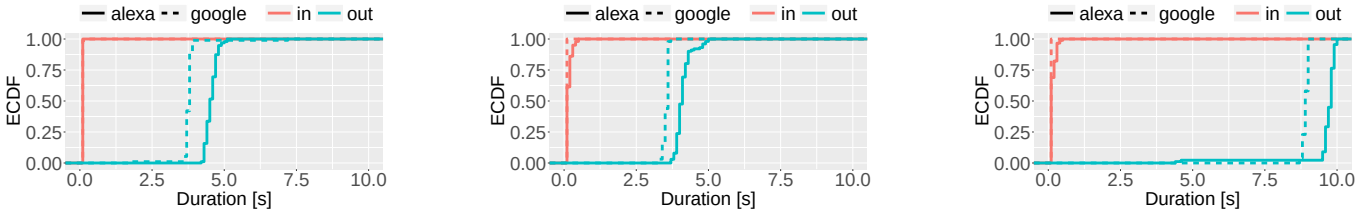
Next, the duration of the two different phases of a transaction, the request and the reply phase, as defined by the time difference between the first and last payload packet arrival in each direction is investigated.

Therefore, Figure 3 shows the distribution of the duration for both identified phases for each scenario. The x-axis shows the phase duration in seconds, the y-axis the empirical cumulative distribution function (ECDF). The linetype indicates the evaluated device and the transmission direction is identified by different colors. In all three evaluated scenarios, the reply phase is nearly identical. The reply is in nearly all cases downloaded in a single time slot of 0.1 seconds. This is explained by the fact that replies contain less than 100 kB of data, as shown in Figure 2. For the request phase indicated in blue, the results show that in all scenarios Google Home exhibits shorter request phases compared to Amazon Alexa. Since the request phrases are, except for the wake word, identical, this is an indication that Google Home buffers more speech data included in the initial burst. When comparing the request phase duration for the three different scenarios, it can



(a) Short duration request phrase (b) Medium duration request phrase (c) Long duration request phrase

Fig. 2. Cumulative downloaded and uploaded data per transaction for both evaluated devices.



(a) Short duration request phrase (b) Medium duration request phrase (c) Long duration request phrase

Fig. 3. Distribution of the durations for upload and download phases for different request durations and devices.

be seen that for the short and medium duration request phrases, no significant difference is observed, although the medium duration request phrase is roughly 2 s longer than the short duration request phrase. This is explained by the fact that the device keeps listening for a minimum time, independent on the requesting phase duration. Thus, it immediately stops listening in the medium and long duration scenarios. We assume that the device waits for further details in the short duration scenario, which triggers the extended listening period. However, this assumption requires further validation. Otherwise, the observed durations correlate with the duration of the request phrase.

Based on the same timestamps used for the evaluation of the phase duration distribution, we investigate the distribution of the processing delay as defined the time difference between the end of the request and the beginning of the reply phase. To this end, Figure 4 shows the corresponding ECDFs for each scenario. The x-axis shows the processing delay in seconds while the y-axis shows the ECDF. The linetype identifies the different devices. The first observation made here are negative instances of the delay in all three scenarios. These negative values are explained by the fact that in some cases, the reply was received before the request phase was completed. Hence the reply phase starts before the request phase ends, which leads to the negative values. One example of this behavior can be seen in Figure 1a. This behavior occurs mainly for Amazon Alexa, and especially in the medium and long duration request scenarios, as we assume that the information required to compute the result is sufficient even before the request phrase is complete. Furthermore, the request phrase has no impact on the measured delay for the Amazon Alexa, while the delay increases significantly for Google Home. As the cloud component of the assistants are largely unknown, we can unfortunately not infer what exactly causes this increase in processing delay. It seems that an increase in voice data leads to increased processing times for the Google Home assistant.

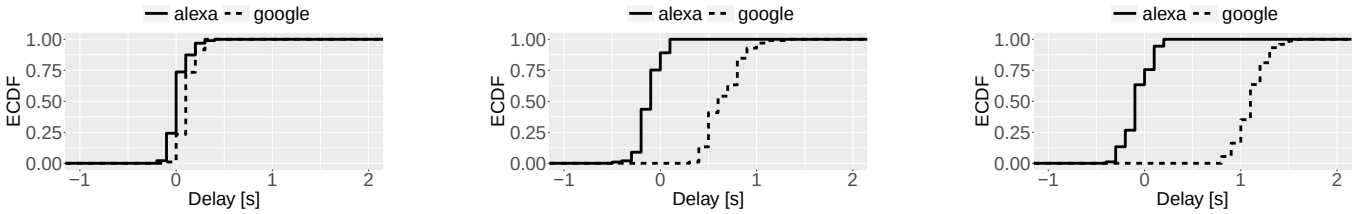
D. Initial Burst Size

Finally, the distribution of the burst size, that occurs at the beginning of the request phase due to the pre-buffered voice data is evaluated. The results, presented in Figure 5 show that the initial burst size for Amazon Alexa remains similar for all three request phrases, which is to be expected, since all phrases are longer than the duration that is pre-buffered by the device. However, the burst size distribution for Google Home shows a slight trend towards larger burst sizes for longer request phrases, which can again be explained by the increased bitrate when it comes to voice data in Google Home.

V. Conclusion

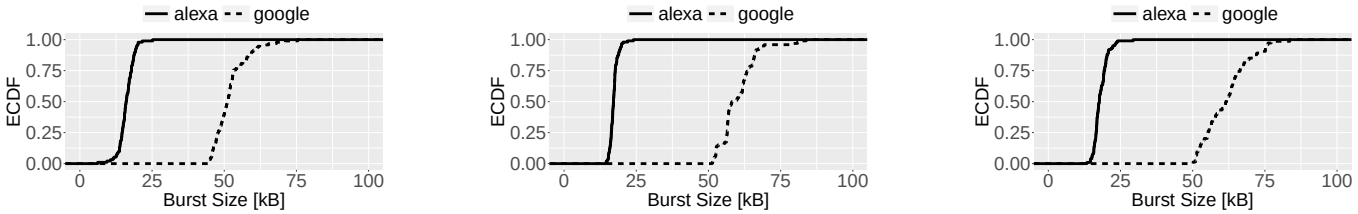
In this work, we present the fundamental background regarding the traffic characteristics of smart home assistants like Amazon Echo Dot and Google Home. The main contribution is a detailed measurement study of the network traffic and requesting behavior of the devices. To this end, we created a dedicated testbed to perform measurements with as little impact on the system behavior as possible while still enabling automated and reproducible measurements. Based on this testbed, we defined various measurement scenarios to evaluate the impact of different parameters on the system behavior.

The general behavior of the system shows that, except for a deterministic heartbeat, both devices only send data to the cloud when addressed directly. Furthermore, we detected similar uplink traffic patterns for both devices, independent from the requested content. The duration and size of request patterns only depends on the request duration. Regarding the processing of a user request, both devices feature similar



(a) Short duration request phrase (b) Medium duration request phrase (c) Long duration request phrase

Fig. 4. Processing duration distribution for different request durations and devices.



(a) Short duration request phrase (b) Medium duration request phrase (c) Long duration request phrase

Fig. 5. Burst size distribution for outgoing data streams.

patterns consisting of an initial burst, followed by an upload phase and finally exhibit a downlink burst to receive the reply from their respective cloud service components.

The main difference between the devices is the request size. The Google Home consistently generated significantly more network traffic and was in general more sensitive to the duration of the user's request phrase compared to the Amazon Echo Dot. From a quality of service perspective, the Google Home buffers more information that is sent to the server as a burst. Furthermore, the Amazon Alexa tends to have a shorter processing delay.

These findings are, together with device usage statistics, beneficial when it comes to developing traffic and usage models as well as evaluating the service quality of home assistants. The main input parameters are the device type, the requesting frequency, and duration. Based on our measurements, the requested content has no influence on the traffic generation in the uplink. Nevertheless, this must be studied in more detail with less commonly used requests.

References

[1] A. Mutchler, "A Timeline of Voice Assistant and Smart Speaker Technology From 1961 to Today," March 2018. [Online]. Available: https://voicebot.ai/2018/03/28/timeline-voice-assistant-smart-speaker-technology-1961-today/

[2] "Smart Speakers To Grow 60% In 2018." January 2018. [Online]. Available: http://www.insideradio.com/free/forecast-smart-speakers-to-grow-in/article_7708397a-f51b-11e7-9f1e-178bfa1db134.html

[3] B. Kinsella, "U.S. Smart Speaker Market Share: Apple Debuts at 4.1%, Amazon Falls 10 Points and Google Rises," April 2018. [Online]. Available: https://voicebot.ai/2018/04/02/smart-speaker-owners-use-voice-assistants-nearly-3-times-per-day/

[4] "Google beats Amazon to first place in smart speaker market," May 2018. [Online]. Available: https://www.canalys.com/newsroom/google-beats-amazon-to-first-place-in-smart-speaker-market

[5] G. Kenny, "I Know Everything About You! The Rise of the Intelligent Personal Assistant," August 2015. [Online]. Available: https://securityintelligence.com/i-know-everything-about-you-the-rise-of-the-intelligent-personal-assistant/

[6] A. Sivanathan, H. H. Gharakheili, F. Loi, A. Radford, C. Wijenayake, A. Vishwanath, and V. Sivaraman, "Classifying iot devices in smart environments using network traffic characteristics," *IEEE Transactions on Mobile Computing*, 2018.

[7] A. Ikpehai, B. Adebisi, and K. Anoh, "Effects of traffic characteristics on energy consumption of iot end devices in smart city," in *2018 Global Information Infrastructure and Networking Symposium*. IEEE, 2018.

[8] A. Sivanathan, D. Sherratt, H. H. Gharakheili, A. Radford, C. Wijenayake, A. Vishwanath, and V. Sivaraman, "Characterizing and classifying iot traffic in smart cities and campuses," in *2017 IEEE Conference on Computer Communications Workshops*. IEEE, 2017.

[9] M. H. Mazhar and Z. Shafiq, "Characterizing smart home iot traffic in the wild," *arXiv preprint arXiv:2001.08288*, 2020.

[10] A. Purington, J. G. Taft, S. Sannon, N. N. Bazarova, and S. H. Taylor, "Alexa is my new bff: social roles, user satisfaction, and personification of the amazon echo," in *CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2017.

[11] B. R. Cowan, N. Pantidi, D. Coyle, K. Morrissey, P. Clarke, S. Al-Shehri, D. Earley, and N. Bandeira, "What can i help you with?: infrequent users' experiences of intelligent personal assistants," in *19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 2017.

[12] M. B. Hoy, "Alexa, siri, cortana, and more: An introduction to voice assistants," *Medical reference services quarterly*, 2018.

[13] F. Bentley, C. Luvogt, M. Silverman, R. Wirasinghe, B. White, and D. Lottridge, "Understanding the long-term use of smart speaker assistants," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2018.

[14] B. Kinsella, "Smart speaker owners use voice assistants nearly 3 times per day," June 2018. [Online]. Available: https://voicebot.ai/2018/06/03/u-s-smart-speaker-market-share-apple-debuts-at-4-1-amazon-falls-10-points-and-google-rises/

[15] F. Loh, V. Vomhoff, F. Wamser, F. Metzger, and T. Hoßfeld, "Traffic measurement study on video streaming with the amazon echo show," in *Proceedings of the 4th Internet-QoE Workshop on QoE-based Analysis and Management of Data Communication Networks*. ACM, 2019.

[16] A. Schwind, F. Wamser, T. Gensler, P. Tran-Gia, M. Seufert, and P. Casas, "Streaming characteristics of spotify sessions," in *10th International Conference on Quality of Multimedia Experience*. IEEE, 2018.

[17] "Speech synthesis markup language (ssml) reference." [Online]. Available: https://developer.amazon.com/docs/custom-skills/speech-synthesis-markup-language-ssml-reference.html

[18] "Ssml — actions on google — google developers." [Online]. Available: https://developers.google.com/actions/reference/ssml