Personal and Ubiquitous Computing (2019) 23:67–79
https://doi.org/10.1007/s00779-018-1174-x

ORIGINAL ARTICLE

# Alexa, are you listening to me? An analysis of Alexa voice service network traffic

Marcia Ford[1] · William Palmer[1]

Received: 9 January 2018 / Accepted: 18 June 2018 / Published online: 28 June 2018


**Abstract**
With approximately 8.2 million Echo family devices sold since 2014, Amazon controls 70% of the intelligent personal assistant market. Amazon's Alexa Voice Service (AVS) provides voice control services for Amazon's Echo product line and various home automation devices such as thermostats and security cameras. In November 2017, Amazon expanded Alexa services into the business intelligent assistant market with Alexa for Business. As corporations integrate Alexa into their corporate networks, it is important that information technology security stakeholders understand Alexa's audio streaming network behavior in order to properly implement security countermeasures and policies. This paper contributes to the intelligent personal assistant knowledge domain by providing insight into Amazon Voice Services behavior by analyzing the network traffic of two Echo Dots over a 21-day period. The Echo Dots were installed in a private residence, and at no time during the experiment did family members or house guests purposely interact with the Echos. All recorded audio commands were inadvertent. Using a k-mean cluster analysis, this study established a quantifiable AVS network signature. Then, by comparing that AVS signature and logged Alexa audio commands to the 21-day network traffic dataset, this study confirmed disabling the Echo's microphone, with the on/off button, prohibits audio recording and streaming to Alexa Voice Service. With 30–38% of Echo Dots' spurious audio recordings were human conversations, these findings support the Echo Dot recorded private home conversations and not all audio recordings are properly logged the Alexa Application. While further Alexa network traffic studies are needed, this study offers a network signature capable of identifying AVS network traffic.

**Keywords** Intelligent personal assistant · Internet of things · Consumer privacy · Privacy policy · Amazon Alexa · Home automation · Amazon voice services

## 1 Introduction

Intelligent personal assistants (IPAs) such as Apple's Siri, Microsoft's Cortana, and Amazon's Alexa perform personal tasks such as managing digital calendar events, creating to-do lists, reporting weather forecast, and streaming music. Amazon controls 70% of the intelligent personal assistant market [1] with approximately 8.2 million Echo family devices sold since its debut in 2014 [2]. Amazon Alexa Voice Service (AVS) provides voice control services for Alexa integrated hardware devices such as, but not limited to, Amazon's Echo family product line and smart home devices such as thermostats and security cameras. In 2017, Amazon announced an LG Alexa-enabled refrigerator [3] and Volkswagen announced it would embed Alexa directly into their automobile systems [4]. Alexa Voice Services are not limited to consumer products. In November 2017, Amazon announced Alexa for Business [5] which is an extension of Alexa Voice Services into the corporate calendar, meeting, conference call, and task management markets. It is predicted 60 million Alexa integrated devices will be sold by 2020 [6] and, with Alexa's current integration with more than 50 individual company products, it appears this prediction aligns with Amazon's strategic vision of "Alexa to be everywhere" [7].

As corporations integrate Alexa into their corporate networks, it is important that information technology security stakeholders understand Alexa's audio streaming network behavior in order to properly implement security countermeasures

✉ Marcia Ford
mcombs@murraystate.edu

William Palmer
wpalmer@murraystate.edu

[1] Institute of Engineering, Murray State University, Murray, KY 42071, USA

and policies pertaining to AVS. Also, understanding AVS audio streaming behavior may alleviate consumer and privacy advocate eavesdropping concerns. Consumers are concerned about personal data privacy, control of personal data lifecycle, granting or withdrawing consent, and understanding the risk of using ubiquitous computing [8]. Consumers' privacy concerns are a barrier to ubiquitous computing's long-term success [9] and could be an obstacle to IPA adoption.

Little is documented about AVS and Echo audio streaming network behavior. Previous Alexa/Echo studies focused on Echo hardware and software [10], Alexa ecosystem artifacts [11], or Echo's internal firewall [12]. One research project [13] attempted to decrypt Alexa Voice Service traffic with a proxy server, but with little success due to Echo's rejection of the proxy server key material.

Researchers at Princeton University identified AVS network patterns by analyzing AVS network traffic while repeating voice commands to an Echo device [14]. Building upon their work, this paper contributes to IPA knowledge domain by providing insight into AVS and Echo audio streaming network behavior by collecting and analyzing all network traffic and logged audio streams of two different Echo Dots. It was expected for every logged audio stream there was an associated AVS network traffic signature. This study attempted to answer three consumer privacy questions: Does the Echo record and stream private conversations to AVS without utilizing the wake word? Are Alexa streamed user commands properly logged and available for review by the consumer? Does the Echo's microphone on/off button completely disable the microphone feature?

## 2 Alexa system architecture

Alexa is a cloud-based software program that performs the duties of an intelligent personal assistant while the Amazon's Echo product line (Echo Dot, Echo Tap, or Echo Show) is the Internet-connected hardware (speaker, microphone, and camera) always "listening" for the user's activation phrase known as the wake word. As shown in Fig. 1, when a user says "Alexa, what time is it?", the wake word "Alexa" signals the Echo to record and stream the subsequent voice commands to Alexa Voice Services. Using an application program interface (API), AVS then passes the voice commands to Alexa which uses speech recognition and natural language understanding for the processing and formulation of Alexa's responses. Alexa, then, responds to the user's commands with Echo's built-in speakers, "The time is 3:15 pm," and sends a copy of the voice command and Alexa's response to the Alexa application located in the Amazon cloud. The copy of the user's audio commands and Alexa's responses are known as response cards. The response cards are accessible via the smartphone Alexa application or directly from the unofficial Alexa API website http://pitangui.amazon.com.

## 3 Experiment design, data collection, and methodology

The purpose of this study was to provide insight into AVS and Echo network behavior by analyzing the network traffic and logged response cards of two Echo Dots and answering the following questions:

- Does the Echo's microphone on/off button completely disable the microphone feature?
- Does the Echo record and stream private conversations to AVS without utilizing the wake word?
- Are Alexa streamed user commands properly logged and available for review by the consumer?

Two identical Amazon second-generation Echo Dots (model #RS03QR and software version 564196920) were installed in the kitchen/family area of the researcher's home. The Echo Dot configurations were identical with the exception of the microphone settings. Echo Dot 1's microphone was enabled using the hardware on/off button while Echo Dot 2's microphone was disabled. Neither Echo Dot underwent voice recognition training or used third-party skills. Response card logging for both Echo Dots was configured to use the researcher's personal Amazon account. Family members and house guests were instructed not to verbally interact with the Echo Dots or change the microphone settings. Both Echo Dots were configured to use the wake word "Echo." As shown in Fig. 2, the Echo Dots connected wirelessly to the Internet through a home router. Charter Spectrum provided Internet access, network address translation, and domain name services (DNS). Firewall services and dynamic host control protocol (DHCP) were provided by a Cisco Linksys WRT 1200AC router. A Ubiquity Nano Station LocoM5, configured as a bridge, provided wireless local area network (WLAN) connectivity for both Echo Dots. A desktop computer running the Wireshark protocol analyzer collected network traffic.

Using Wireshark, an open-source protocol analyzer, all Echo Dot ingress and egress network traffic was collected for 21 days, 24 h a day from February 27, 2017, through March 19, 2017, with Wireshark's command line utilities dumpcap.exe, mergecap.exe, and tshark.exe were used to read, write, and export network traffic into a comma-separated value (CSV) file. The Wireshark CSV column attributes exported were frame time, IP source address, IP destination address, TCP source port, TCP destination port, protocol, TCP length, TCP stream index, and information column. Echo Dot network traffic datasets were created by importing Wireshark's CSV files into Tableau Desktop 10.5 and Microsoft PowerPivot.

This study attempted to answer the three research questions by finding inconsistencies between the number of inadvertent

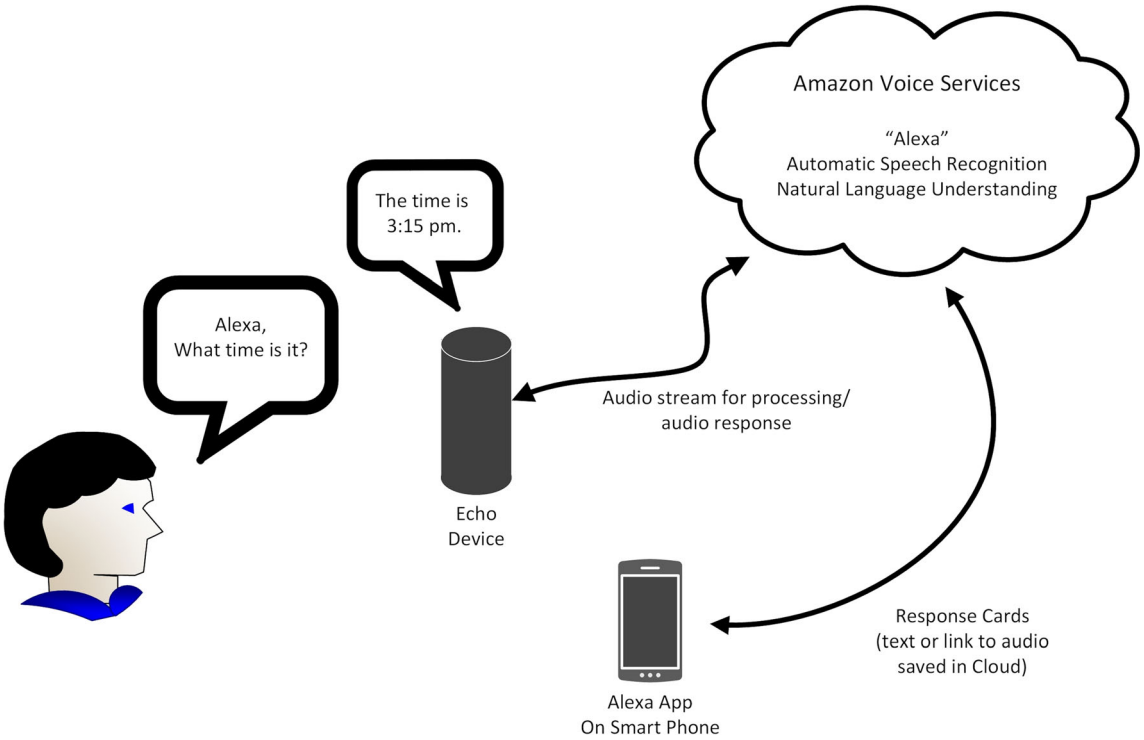


**Fig. 1** Alexa system architecture

logged response cards and AVS audio command network traffic. Enumeration of the logged response cards meant reading, dating, and transcribing all the logged response cards from the smartphone Alexa application. However, a quantifiable AVS network signature was necessary to identify AVS audio commands within network traffic. The AVS network signature was determined by running Tableau's k-mean clustering algorithm using all the AVS network traffic data points.

The k-mean clustering method was used to group records when a set of input fields were available to group records based on similarity of values in order to discover k clusters. The records within each cluster are similar while they are distinct from the records in other clusters. Tableau 10.5 k-mean clustering analytic tool uses the hierarchical divisive method, Howard-Harris, to determine initial centroids, and then identifies the k-mean clusters with Lloyd's algorithm with squared Euclidean distances. Since the value $k$ was not specified, Tableau determines the optimal number of clusters using the Calinski-Harabasz index [15, 16].

To answer the research questions, this study needed to locate inconsistencies between Alexa application logged response cards and identified AVS network traffic, this was accomplished by comparing daily logged response cards to AVS network traffic. To assess the robustness of the methodology, three different network traffic datasets were collected and evaluated. The next two sections discuss the documentation of response cards and identification of the AVS network signature.

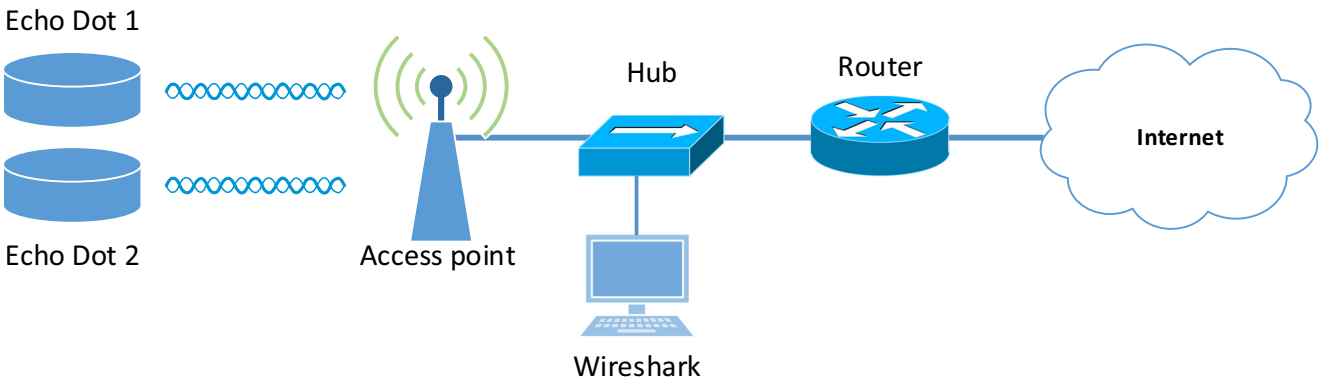


**Fig. 2** Home network topology

### 3.1 Documenting Alexa application response cards

The copy of voice commands recorded and streamed by Echo to AVS for natural language processing is known as a response card and is accessible from the Alexa application history settings. As shown in Fig. 3, a response card contains the transcribed text of the streamed audio (echo morning), the AVS's response (Good morning, and happy birthday….), and an optional play button linking to the actual voice recording stored in the Amazon Web Service (AWS) cloud. The response card also provides user verification (Yes or No) to the accuracy of AVS interpreted command.

At no time during the experiment did humans purposely interact with the Echo Dots with the wake word "Echo". All audio recorded and streamed by the Echo Dot(s) were inadvertent, meaning the Echo misinterpreted home audio as the wake word, or it randomly recorded home audio. Upon review of the researcher's smartphone Alexa application history settings for February 27, 2017, through March 19, 2017, Echo Dot 1 (ED1) with the microphone enabled logged 33 response cards while Echo Dot 2 (ED2) with the microphone disabled logged 0 response cards. Thirteen of ED1's response cards pointed to previously recorded audio; therefore, they were classified as duplicate response cards. For example, on March 14, a total of three response cards were logged with different transcribed text, but the play button (link) pointed to the same audio of a television commercial. Therefore, those response cards were logged as 1 original and 2 duplicates. It was assumed, duplicate response cards were created by multiple interactions between AVS cloud servers and Alexa application and not by network retransmits from Echo to AVS. Therefore, duplicate response cards were not considered for data analysis purposes.

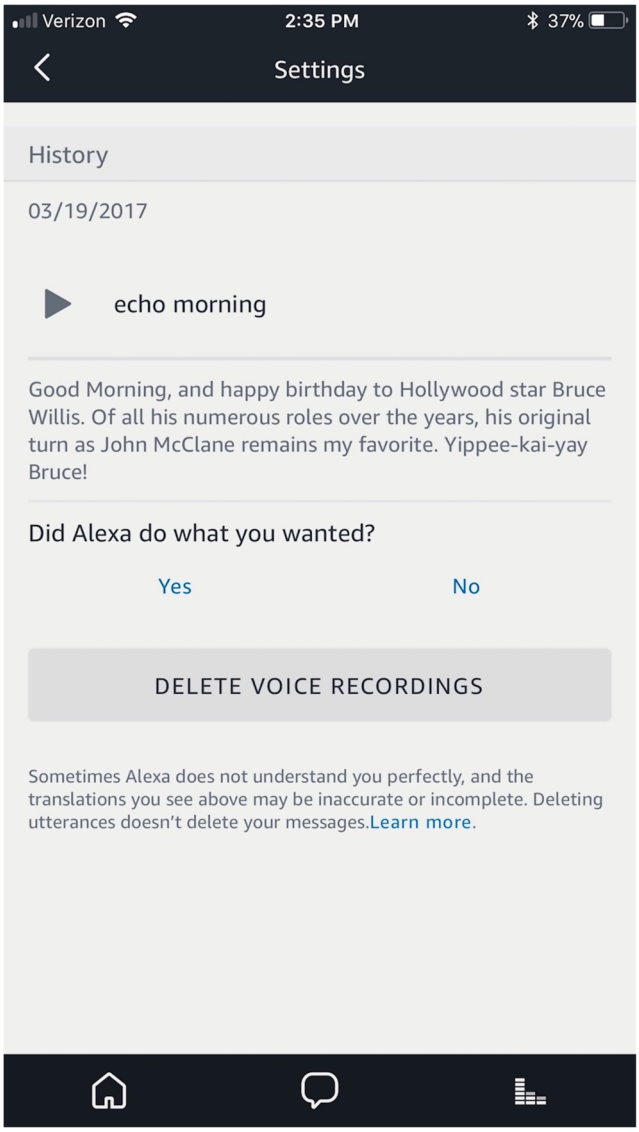


**Fig. 3** Image of ED1 Alexa Application smartphone response card

The Alexa iOS application version used during this study was capable of logging only the response card creation date but not creation time. Therefore, only response card creation dates, text, and audio were logged. Another limitation of the iOS Alexa application was when reviewing response cards, it was necessary to periodically close and reopen the application to prevent response cards from disappearing and reappearing. The response cards shown in Table 1 were verified by completing three iterations of the response card documentation process. As shown in Table 1, ED1 (mic on) logged 20 original and 13 duplicate response cards, whereas ED2 (mic off) logged 0 response cards. Of all the inadvertent household recordings, based on the logged response cards, 70% of the recordings were TV audio and 30% human voice. A little less than half (40%) of the total logged response cards were duplicates.

### 3.2 Identification of AVS network signature

Wireshark analysis showed the majority of egress Echo Dot traffic was encrypted with Transport Layer Security (TLS) also known as Secure Socket Layer (SSL). Decrypting Echo Dot AVS traffic for analysis was unlikely since Amazon cryptographic key material is unavailable. The decryption of SSL traffic with a proxy server was not possible since the Echo operating system rejects foreign server SSL certificates [13]. Therefore, the volume of egress AVS encrypted traffic was measured by summing SSL record type 23 (application data) Transport Control Protocol segment lengths (tcp.len). It was assumed SSL transport control segment lengths would increase with the presence of streamed audio.

It was expected when a user asked Alexa a question or issued a voice command, Echo Dot egress AVS network traffic volume would increase, and as the transmission concluded, the AVS network traffic volume would decrease. The AVS signature was established by collecting Echo Dot network

**Table 1** Number of daily response cards and duplicates logged by Echo Dot(s) for February 27–March 19, 2017 (days with no response cards are omitted)

| Response card creation date | Echo device | Total number of response cards | Number of original response cards | Number of duplicate response cards | Original response card type (TV or human) |
|---|---|---|---|---|---|
| Feb 27 | Echo Dot 1 | 2 | 1 | 1 | TV (1) |
| Feb 28 | Echo Dot 1 | 3 | 2 | 1 | Human (2) |
| Mar 1 | Echo Dot 1 | 1 | 1 | 0 | TV (1) |
| Mar 2 | Echo Dot 1 | 1 | 1 | 0 | TV (1) |
| Mar 3 | Echo Dot 1 | 5 | 4 | 1 | Human (4) |
| Mar 4 | Echo Dot 1 | 1 | 1 | 0 | TV (1) |
| Mar 10 | Echo Dot 1 | 2 | 1 | 1 | TV (1) |
| Mar 11 | Echo Dot 1 | 1 | 1 | 0 | TV (1) |
| Mar 13 | Echo Dot 1 | 2 | 2 | 0 | TV (2) |
| Mar 14 | Echo Dot 1 | 3 | 1 | 2 | TV (1) |
| Mar 15 | Echo Dot 1 | 6 | 2 | 4 | TV (2) |
| Mar 19 | Echo Dot 1 | 6 | 3 | 3 | TV (3) |
| | Total | 33 | 20 | 13 | TV (14) Human (6) |

traffic while asking Alexa four different questions. Wireshark analysis of this traffic revealed a sharp increase in SSL traffic to the pindorama.amazon.com domain when asking Alexa questions. A time series analysis of this network traffic dataset revealed, as each question was asked, the sum (tcp.len) of application data to the pindorama.amazon.com domain significantly increased. As shown in Fig. 4, ED1's sum (tcp.len) of application data experienced spikes of approximately 23 k, 13 k, 17 k, and 23 k bytes per second (bps) with each spike associated with an asked question. This pattern of application data spikes was comparable to other established AVS traffic analysis studies [14].

As shown in Fig. 4, in addition to the audio spikes, the egress pindomrama traffic included three instances of 122 bps traffic plus a single instance of 1539 bps that did not appear to be associated with the audio spikes. This suggested Echo Dot AVS traffic included cyclical traffic as well as audio data spikes.

Verification of cyclical AVS traffic was confirmed by analyzing the AVS egress traffic from a time period within the 21-day dataset least likely to contain audio. A time series plot of both Echo Dot sum (tcp.len) application data bytes per minute from March 5, 2017 02:00-02:01 to pindorama.amazon.com, shown in Fig. 5, displayed a flat, repetitive byte pattern of 41 and 82 bytes per minute (bpm) with sporadic transmissions of 192–198 and 3497–3501 bpm. This suggested both Echo Dots transmitted two or three cyclical traffic patterns to the pindorama.amazon.com, and it was suspected that this traffic was a type of heartbeat or keep-alive traffic between the Echo Dots and the pindorama.amazon.com domain. Since the network traffic was encrypted, it was unclear of this traffic's purpose or if it contained any audio transmissions from the Echo Dots. For the purpose of this paper, the AVS cyclical traffic is referred to as keep-alive traffic.

With the identification of AVS audio spikes and keep-alive traffic, a byte per minute fingerprint or signature was needed to distinguished between the two AVS traffic types. Using the larger 21-day dataset, AVS egress traffic was grouped into five clusters based on the sum(tcp.len) variable using a k-mean cluster analysis. Tableau 10.5 k-mean clustering analytic tool uses the hierarchical divisive method, Howard-Harris, to determine initial centroids, and then identifies the k-mean clusters with Lloyd's algorithm with squared Euclidean distances. Since the value $k$ was not specified, Tableau determines the optimal number of clusters using the Calinski-Harabasz index [15, 16].

It was expected the k-mean cluster analysis would identify keep-alive and AVS audio traffic groupings with centers similar to data points identified in Figs. 4 and 5. It was posited keep-alive clusters would not be affected by the microphone settings therefore, ED1 and ED2 keep-alive clusters would be similar in the number of items and data point maximum, minimum, and centroid values. Due to the repetitive behavior of keep-alive traffic, these clusters would also contain a high number of ED1 and ED2 paired data points with lower sums (tcp.len) of application data. It was also expected the k-mean cluster analysis would group AVS audio clusters with minimum data points with a higher sum (tcp.len) application data values and since ED2's microphone was disable, these clusters would not contain any data points from ED2.

The k-mean clustering analysis results grouped the Echo Dot(s) sum(tcp.len) application data into five clusters, as shown in Table 2. Cluster 1, with 96% of the data points, contained items with low center, minimum, and maximum values and was fairly distributed between the Echo Dots. Therefore, Cluster 1 was labeled as a keep-alive cluster. While Cluster 2's center, minimum, and maximum values

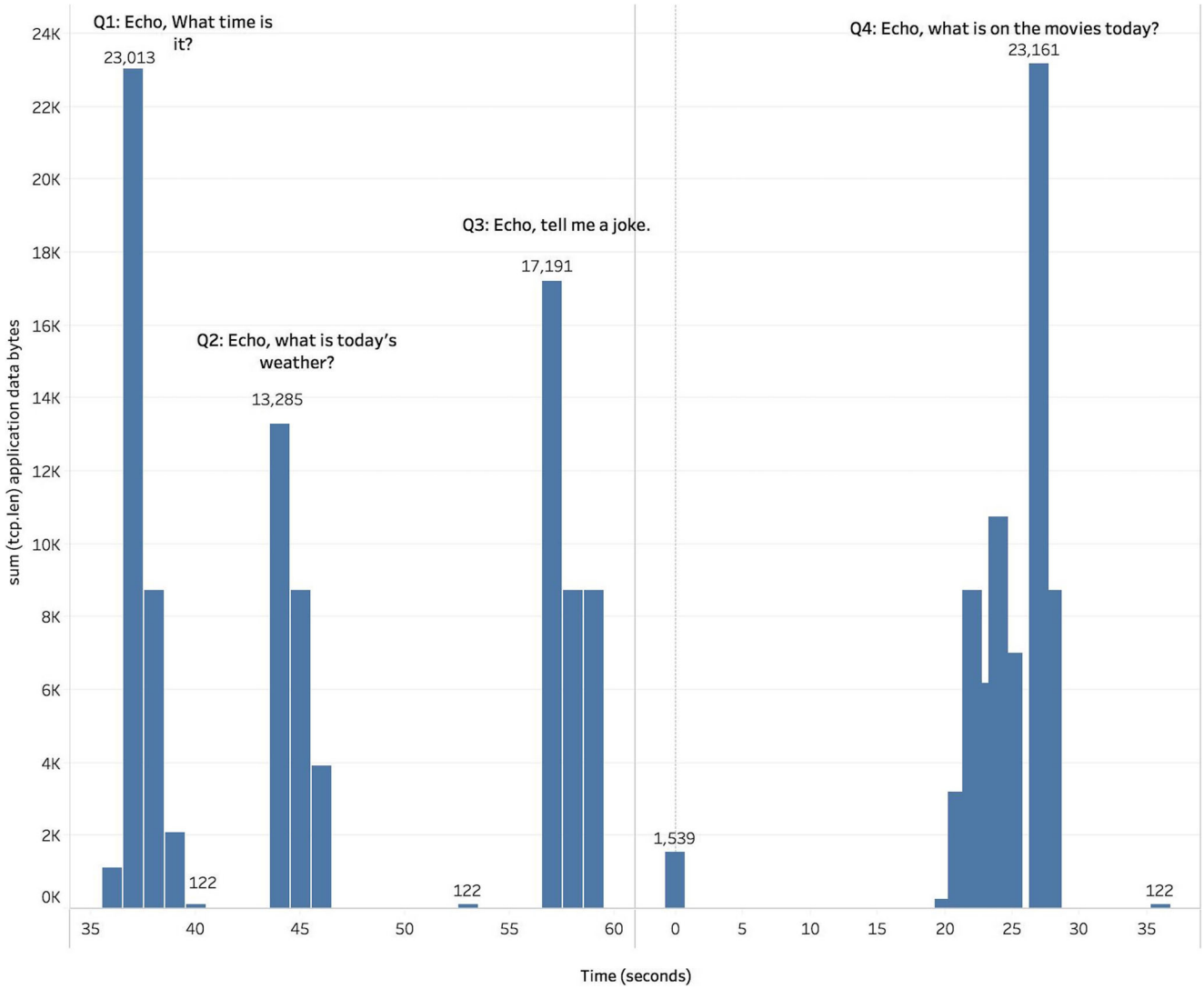


**Fig. 4** Time series plot of Echo Dot 1 sum (tcp.len) of application data bytes per second to the pindorma.amazon.com domain while asking Alexa questions

had increased, ED1 had twice as many data points as ED2. Results indicated that Cluster 2 data points were affected by the microphone settings and could contain audio transmissions.

A whisker-and-box plot of Cluster 2 data points as shown in Fig. 6, reveals the difference in ED1 and ED2's data point spread. With several of ED1's data points close to the lower limits of Cluster 3, it was unclear if these segments contained audio transmissions.

Since Clusters 3–5 did not contain any ED2 data points and the number of ED1 data items equaled the number of logged original response cards, it was suspected that these clusters contained audio transmissions. With the uncertainty of the presence of audio transmissions in Cluster 2, the AVS network traffic signature was defined as Cluster 2's maximum (tcp.len) value of 7049 bpm. The AVS signature stated, any egress Echo Dot application data with a sum(tcp.len) greater than 7049 bpm destined for the pindorama.amazon.com domain was labeled as AVS audio traffic.

In the next section, the defined AVS signature and logged response cards creation dates were compared to the larger 21-day network traffic dataset for the purpose of identifying undocumented AVS audio commands.

## 4 Findings

This section is divided into three parts: Comparison of Response Cards to AVS Traffic, AVS network traffic signature reliability, and Other Echo Dot network traffic.

Comparison of response cards to AVS traffic

Within the 21-day network traffic dataset, it was expected for every logged response card that there was an associated AVS application data point greater than 7050 bpm. It was also

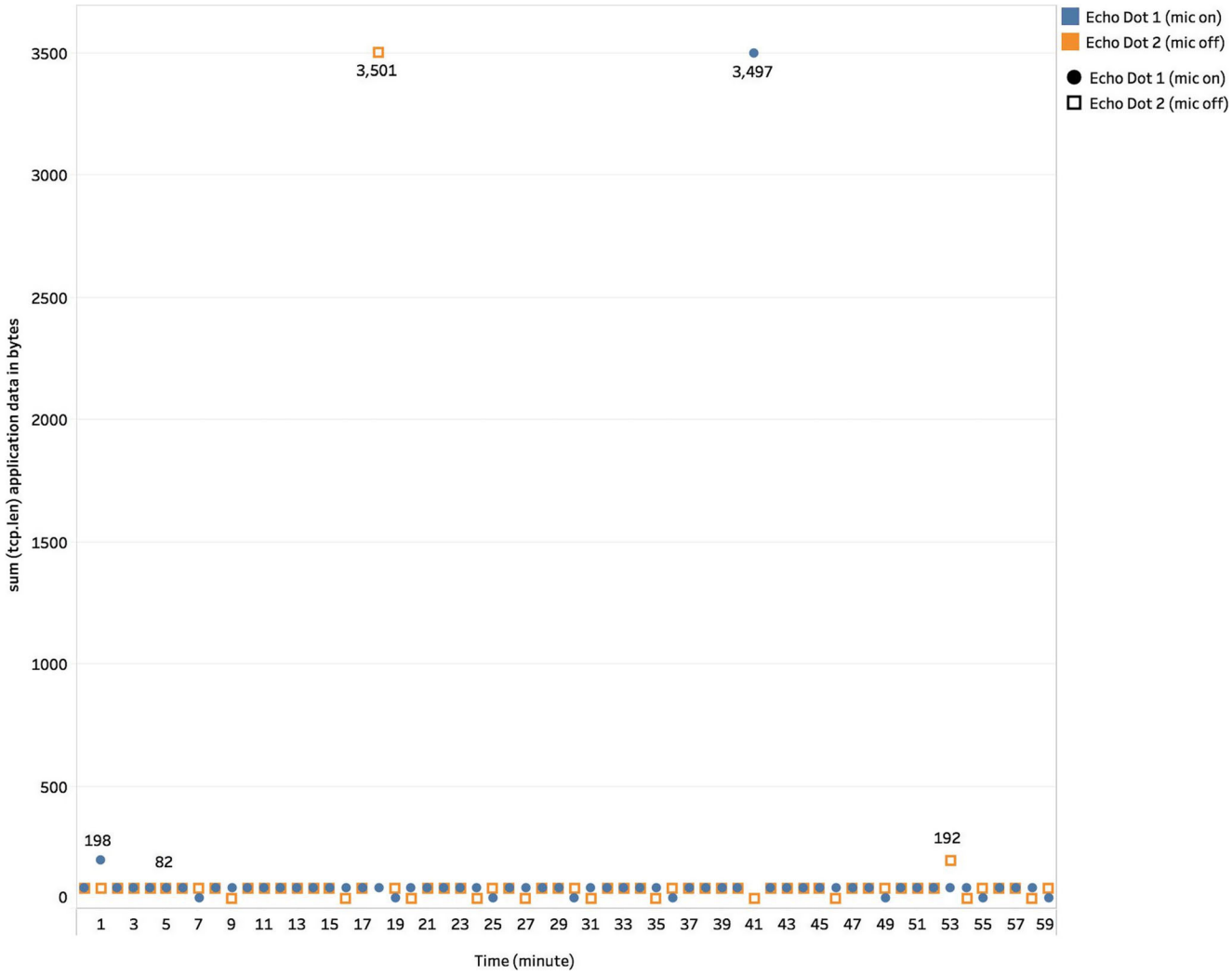


**Fig. 5** Time series plot of Echo Dot(s) sum (tcp.len) application data bytes per minute to the pindorma.amazon.com domain March 5, 2017 02:00

expected that ED2 (mic off) would not log any response cards or transmit any sum(tcp.len) data points to pindorama.amazon.com above 7050 bpm.

Comparing the creation dates and number of original response cards, shown in Table 3, to the AVS data points greater than 7050 bpm, shown in Fig. 7, revealed that for every ED1 logged response card, there was a corresponding AVS data point greater than 7050 bpm. For instance, on February 27, ED1 recorded one original response card that was corresponded with one data point above 7050 bpm. Again, on February 28, ED1 logged two original response cards that corresponded to two data points above 7050 bpm. Echo Dot 2

**Table 2** 21-day dataset k-mean clustering results with application data sum (tcp.len) in bytes per minute

| | Number of items | ED1 Number of items | ED2 Number of items | Centers (sum tcp.len) | Minimum value | Maximum value |
|---|---|---|---|---|---|---|
| Cluster 1 | 58,001 | 28,538 | 29,463 | 79.664 | 41 | 1671 |
| Cluster 2 | 2128 | 1421 | 707 | 3503.8 | 1879 | 7049 |
| Cluster 3 | 6 | 6 | 0 | 31,302.0 | 24,084 | 45,804 |
| Cluster 4 | 11 | 11 | 0 | 61,444.0 | 49,668 | 71,364 |
| Cluster 5 | 3 | 3 | 0 | 87,393.0 | 81,308 | 99,415 |
| Not clustered | 0 | 0 | 0 | | | |
| Total | 60,149 | | | | | |

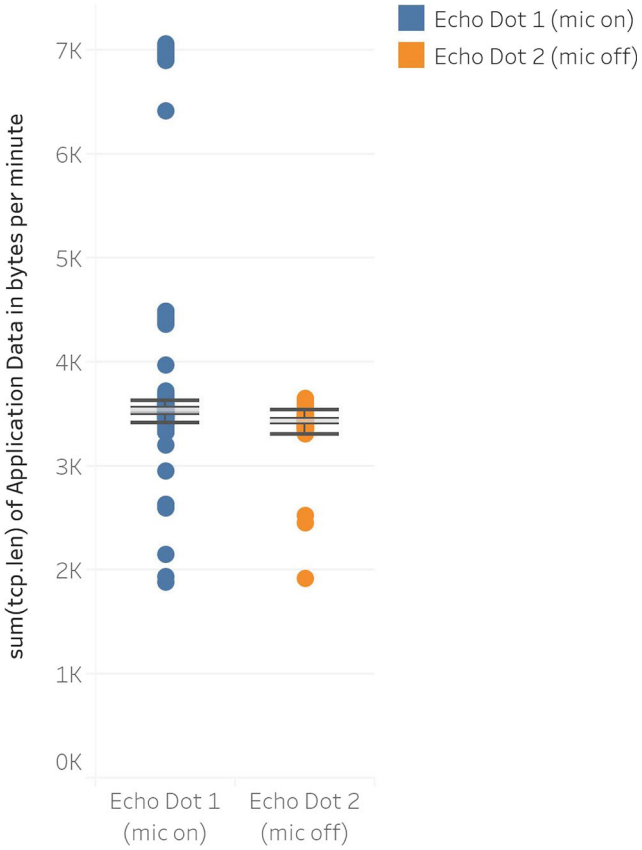


**Fig. 6** Box-and-whisker plot of k-mean Cluster 2 results

(mic off) did not log any response cards or experience any sum (tcp.len) application data spikes greater than 7050 bpm. ED2 only transmitted keep-alive traffic to the pindorma.amazon.com domain as shown earlier in Fig. 5.

Two inconsistencies were identified when pairing the response cards to AVS data points. On March 14, 2017, ED1 logged one original response card with no associated data point greater than 7050 bpm. This was most likely due to a 5-min power outage on March 14 at 18:50 CST causing loss of network traffic. But if the response cards are created after AVS receives the audio stream, a power outage should have caused the loss of audio stream and loss of logged response cards, not just audio stream. For that reason, it is unclear if the power outage caused the logged response card to data point inconsistency on March 14, 2017.

The second inconsistency was on March 15, where ED1 logged six response cards (two original and four duplicates) as shown in Table 4, but three AVS data points greater than 7050 bpm were identified instead of the two expected data points. This evidence suggests the extra AVS data point was not part of the response card but a separate AVS data point. It is unclear what caused the extra AVS data point.

In the next section, the reliability of the AVS signature (application data greater than 7050 bmp) was assessed using two different 2018 network traffic datasets that were collected from an Echo Show, Echo Dot(s), Echo (Tall), and a Google Home.

### 4.1 AVS network traffic signature reliability

To assess the reliability of the AVS signature, two additional AVS network traffic analyses were conducted using network traffic from two Echo Dots and from a single Echo Dot, Echo (Tall), Echo Show, and Google Home. Using the same network topology in Fig. 2, the original experiment was repeated creating a 10-day dataset with network traffic from two Echo Dots on April 17–26, 2018. Once again, Echo Dot 2's microphone was disabled. Both Echo Dots were running current

**Table 3** Number of daily response cards logged by Echo Dot(s) during 21-day collection period February 27–March 19, 2017 (days with no response cards are omitted)

| Response card creation date | Echo device | Total number of response cards | Number of original response cards | # of data points > than 7050 bpm |
|---|---|---|---|---|
| Feb 27 | Echo Dot 1 | 2 | 1 | 1 |
| Feb 28 | Echo Dot 1 | 3 | 2 | 2 |
| Mar 1 | Echo Dot 1 | 1 | 1 | 1 |
| Mar 2 | Echo Dot 1 | 1 | 1 | 1 |
| Mar 3 | Echo Dot 1 | 5 | 4 | 4 |
| Mar 4 | Echo Dot 1 | 1 | 1 | 1 |
| Mar 10 | Echo Dot 1 | 2 | 1 | 1 |
| Mar 11 | Echo Dot 1 | 1 | 1 | 1 |
| Mar 13 | Echo Dot 1 | 2 | 2 | 2 |
| Mar 14 | Echo Dot 1 | 3 | 1 | 0 |
| Mar 15 | Echo Dot 1 | 6 | 2 | 3 |
| Mar 19 | Echo Dot 1 | 6 | 3 | 3 |

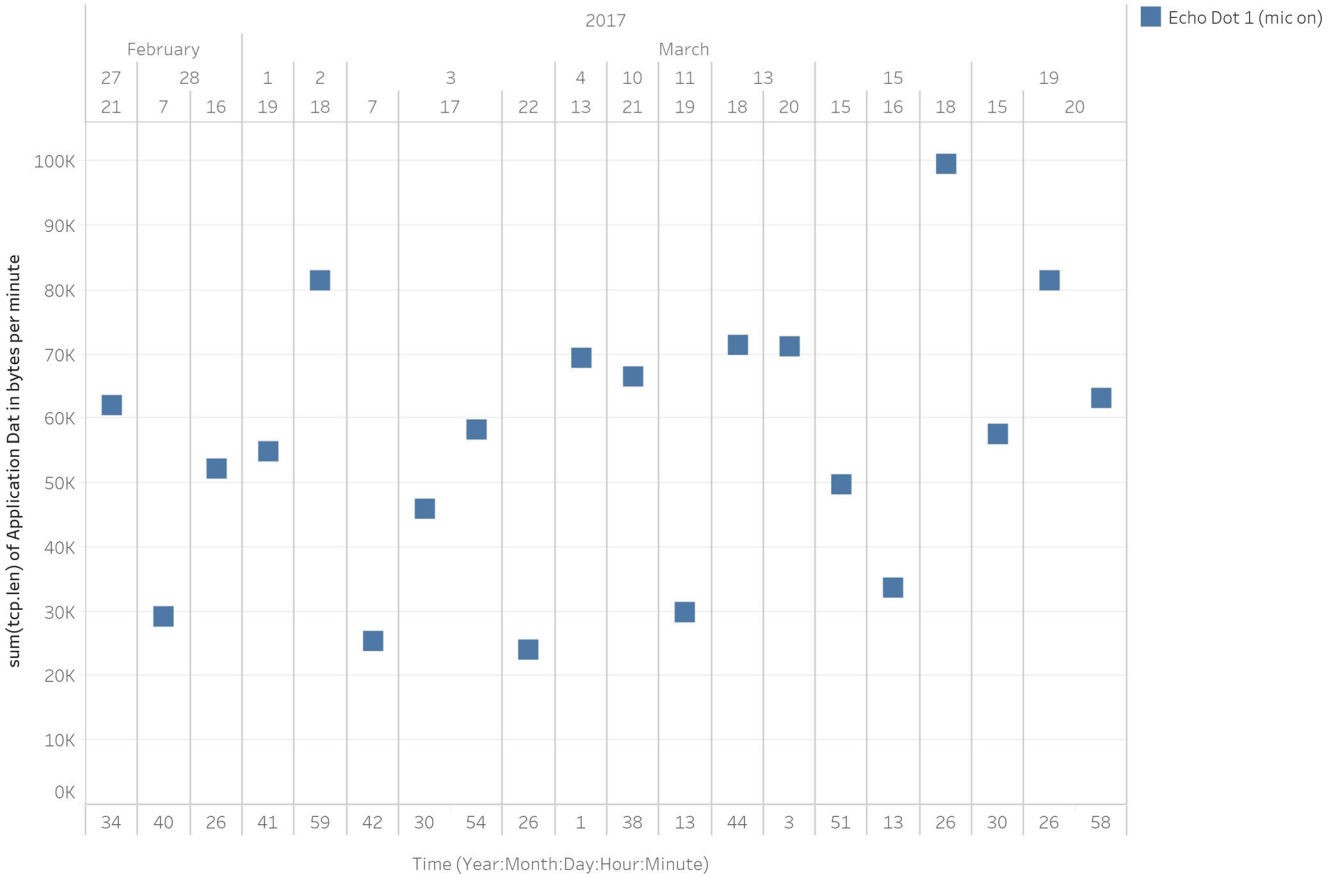


**Fig. 7** Time series plot of Echo Dot(s) egress application data sum(tcp.len) greater than 7,050 bpm to pindorma.amazon.com

firmware version and the Apple iOS device was running the most current version of the Alexa application.

The Echo Dots were placed in same location in researcher's home as the first experiment. They were reset to their default settings and re-registered to a different Amazon account. During the data collection period, no human purposely interacted with the Echo Dots. Alexa response cards were transcribed daily, without issues.

A second 22-minute dataset was created by collecting the network traffic from an Echo Dot, Echo Show, Echo (Tall), and Google Home while asking four different questions. The Echo (Tall) term was used to distinguish between the first-generation Echo and the Echo Dots. Questions were asked in sequence, starting with Google Home, then Echo Dot, Echo (Tall), and lastly Echo Show.

Analysis of the Echo Dot 10-day 2018 dataset revealed two separate AVS signature inconsistencies, on April 17 and April 22. On April 17, Echo Dot 1 logged three original response cards; however, four AVS data points were identified in the network traffic. On April 22, Echo Dot 1 logged 1 response card with no corresponding AVS data point identified in the network traffic. Echo Dot 2 did not log any response cards or create any AVS network traffic. As shown in Table 5, only Echo Dot 1 (mic on) logged response cards, with 61.5% of those recordings being TV audio and 38.4% being human. Thirty-five percent of total response cards were duplicates. The Echo

**Table 4** Alexa application response card transcription for March 15, 2017 (chronological order)

| Text entry | Audio recording transcription |
|---|---|
| - Echo to drive down | TV audio "Then comes…James Bond.." |
| - Text not available Click to play recording | Repeat TV audio "Then comes…James Bond.." |
| - Echo | TV audio "I don't think so" |
| - Text not available Click to play recording | Repeat TV audio "I don't think so" |
| - Text not available Click to play recording | Repeat of TV audio "I don't think so" |
| - Text not available Click to play recording | Repeat of TV audio "I don't think so" |

**Table 5** Number of daily response cards logged by Echo Dot(s) during the 10-day collection period (days with no response cards are omitted)

| Response card creation date | Echo device | Total number of response cards | Number of original response cards | # of data points > than 7050 bpm | Original response card type (TV or human) |
|---|---|---|---|---|---|
| April 17 | Echo Dot 1 | 5 | 3 | 4 | TV (1) H (2) |
| April 18 | Echo Dot 1 | 1 | 1 | 1 | TV (1) |
| April 19 | Echo Dot 1 | 2 | 2 | 1 | TV (2) |
| April 21 | Echo Dot 1 | 4 | 1 | 1 | H (1) |
| April 22 | Echo Dot 1 | 1 | 1 | 0 | H (1) |
| April 23 | Echo Dot 1 | 3 | 2 | 2 | TV (2) |
| April 25 | Echo Dot 1 | 1 | 1 | 1 | TV (1) |
| April 26 | Echo Dot 1 | 3 | 2 | 2 | TV (1) H (1) |
| | Total: | 20 | 13 | 12 | TV (8) H (5) |

Dot 10-day 2018 dataset was absent of ED2 response cards or AVS network traffic, providing additional evidence that disabling the microphone prohibits audio recording and streaming.

Analysis of 22-min April 30, 2018, Echo and Google Home dataset revealed, for each question asked, shown in Table 6, the Echo and Google Home devices had an associated application data point greater than 7050 bpm, as shown in Fig. 8. Each question appeared in the Alexa App settings and Google Home Assistant history. However, for the first question, Echo (Tall) had two data points greater than 7050 bpm. Since all egress application data was captured, the Echo (Tall)'s 47 k bpm data point destination was pindorma.amazon.com while the 10 k bpm data point destination was device-metrics-us.amazon.com.

## 4.2 Other Echo Dot network traffic

During the 2017 data collection period, the Echo Dot(s) transmitted encrypted data to multiple domains other than pindorma.amazon.com. As seen in Fig. 9, the Echo Dots, in the February–March, 2017, dataset, transmitted encrypted data to ten other domains with the highest volume to the device-metrics-us.amazon.com and dp-gw-na.amazon.com. A bar chart of egress Echo Dots sum(tcp.len) application data to Amazon domains less pindorama as shown in Fig. 9 revealed egress network traffic was not distributed evenly between the Echo Dots, suggesting that the microphone setting may have influenced the egress traffic volume. For example, Echo Dot 1 transmitted approximately 149 k bytes to dcape-na.amazon.com while Echo Dot 2 transmitted zero bytes. Interestingly, only Echo Dot 2 contacted the api.amazom.com domain. Due to encryption, it is unknown, what data was transmitted or what role these domains played in the operation and functionality of the Echo Dots, or if this traffic violated any user personal privacy.

Other Echo Dots network traffic included typical protocols such as Domain Name Server (DNS) protocol, Network Time Protocol (NTP), Internet Control Message Protocol (ICMP), Echo Protocol, and the Hyper Text Transfer Protocol (HTTP).

Clear text HTTP offered some insight into the update and probe activity of the Echo Dots. The Echo Dots communicated with HTTP to the five domains listed as follows:

1. amzdigitaldownload.s.edgesuite.net (a1910.d.akamai.net)
2. weblb-wg.dual-gslb.spotify.com (esdk-ffl.spotify.com),
3. example.com
4. kindle-time.amazon.com
5. spectrum.s3.amazonaws.com, s3-directional-w.amazonaws.com, s3-1-w.amazonaws.com

Both Echo Dots updated their firmware from 564196920 version to 571207720 on March 9 and 10, 2017, by issuing a HTTP get to download the firmware file update-kindle-fu.bin from the amzdigitaldownload.s.edgesuite.net domain. A single HTTP fire and forget message to the weblb-wg.dual-gslb.spotify.com (esdk-ffl.spotify.com) domain was sent by Echo Dot 1 and 2 on March 9 and 17, respectively. After the Spotify server returned the HTTP/1.1200 OK, it immediately closed the connection. No further communication between the Echos and Spotify occurred again. The Echo Dots sent multiple HTTP gets to example.com domain, establish a logical connection, and immediately close the connection without transferring any data.

**Table 6** Questions presented to IPA devices on April 30, 2018

| IPA device questions |
|---|
| 1. What time is it in Regensburg, Germany? |
| 2. How do you say goodbye in German? |
| 3. What is the weather in Regensburg, Germany? |
| 4. How do you say "How are you" in German? |

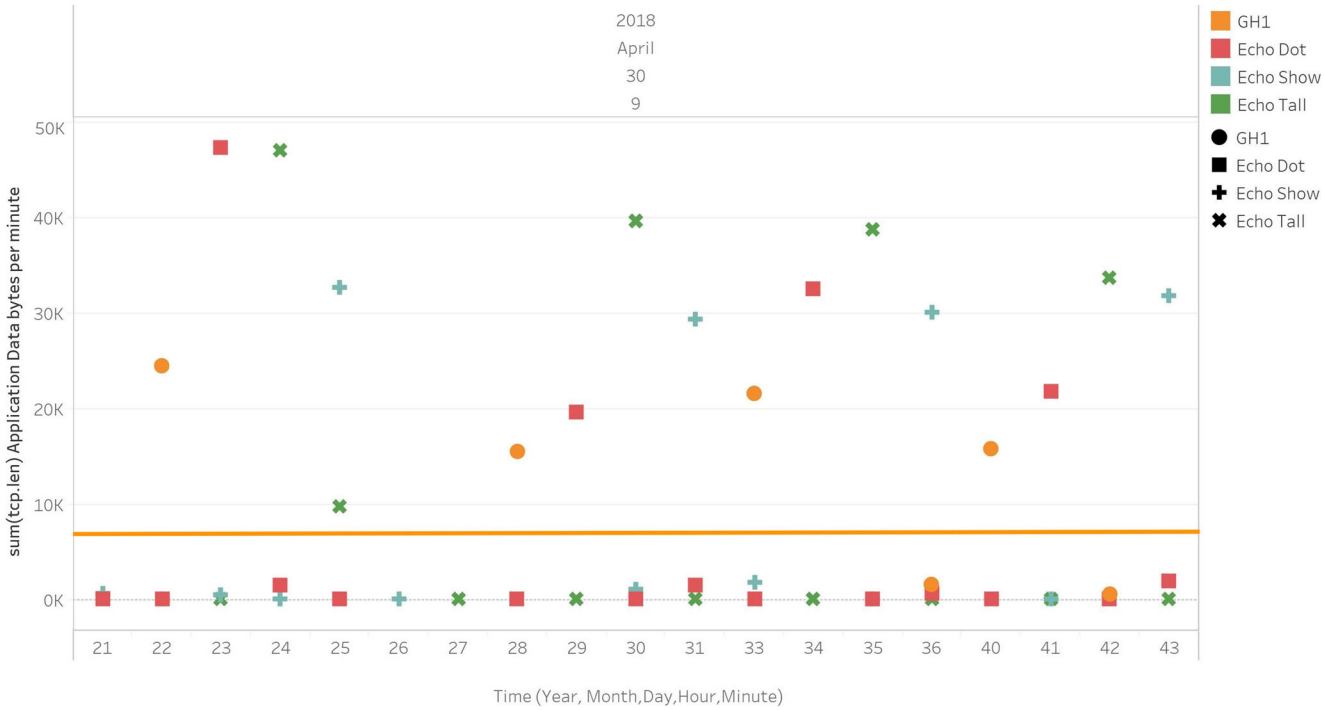


**Fig. 8** Time series plot of Echo and Google Home devices' egress application data while asking four questions April 30, 2018

It was suspected that the HTTP requests to the spectrum and kindle-time domains were Wi-Fi and Internet connectivity probes. The HTTP traffic to these domains were systematic, with the Echos transmitting approximately 400 HTTP requests per day to the spectrum domains and four to five requests to the kindle-time domain. According to Internet mobile forums [17], the HTTP GET/kindle-wifi/wifistub-echo.html request to the host spectrum.s3.amazonaws.com was a reachability probe used by Amazon Kindles. The Echo Dots were sending and receiving the same HTTP traffic described in the Mobile blog, and it was suspected that the Echo Dots utilized the same HTTP probe as the Kindle. It is suspected, after the HTTP GET /kindle-wifi/wifistub-echo.html, the Kindle and Echos expect the spectrum server to return the value 81ce4465-7167-4dcb-835b-dcc9e44c112a. If this value is not returned, the devices sense that they do not have Wi-Fi connectivity. The Echo Dots also sent HTTP request to the kindle-time.amazon.com domain. After the HTTP request, the server immediately closed the connection with a HTTP 403 error and not data transferred.

## 5 Discussion

The purpose of this study was to provide insight into AVS and Echo network behavior by analyzing the network traffic and logged response cards of two Echo Dots in an attempt to answer three questions;

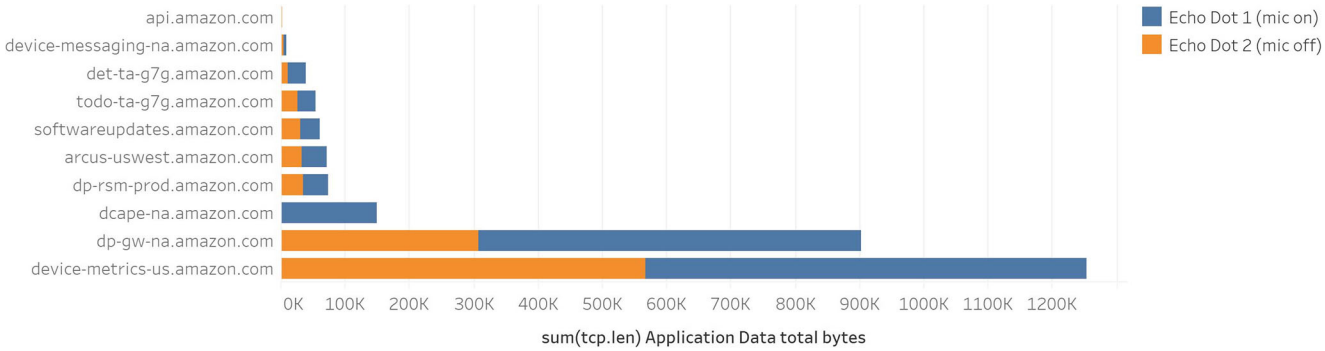


**Fig. 9** Bar Chart of Echo Dot(s) Application Data traffic to Amazon domains less pindorama February–March, 2017

- Does the Echo's microphone on/off button completely disable the microphone feature?
- Does the Echo record and stream private conversations to AVS without utilizing the wake word?
- Are Alexa streamed user commands properly logged and available for review by the consumer?

The absence of Echo Dot 2 (mic off) logged response cards and AVS network traffic greater than 7050 bpm, in two different Echo Dot datasets, established that ED2 did not stream any private conversations, TV audio, or other sounds to Amazon AVS cloud. These results demonstrated when the Echo microphone on/off button is in the off position, the microphone is indeed off, thus answering in the positive to the question Does the Echo's microphone on/off button completely disable the microphone feature?

Alexa users overly concerned about privacy can confidently disable the Echo's microphone and expect that the microphone is not listening. Disabling the microphone limits Alexa's hands-free features such as playing music and controlling the home environment, features for which it was purchased. In regard to Amazon Echo, consumers must find their own balance between privacy and usability.

In this study, during the 21- and 10-day collection periods, no human purposely issued the wake word "echo" or interacted with the Echo Dots. All logged response cards and AVS audio traffic were inadvertent recordings. For the 21- and 10-day Echo Dot datasets, 70 and 61.5% of the logged response cards were TV audio while 30 and 38.4% were human audio, respectively. The presence of human voice response cards demonstrates the Echo Dots recorded home audio without consumers explicitly issuing the wake word. With 30–38% of the Echo spurious recordings consisting of human voices, the answer to research question 2 was answered in the positive, the Echo Dot did record and stream private conversations to AVS without utilizing the wake word.

Since, the Echo Dots did not undergo voice training and were "always listening" for the wake word, it is unclear if Echo Dot was merely attempting to decipher the wake word or nefariously recording private conversations, the answer to that question is unclear. Even if an Echo mistakenly records and streams portions of private conversation, business owners and consumers should understand their conversations are not protected under the United States Constitution 4th Amendment. Due to the long-held legal "third-party doctrine," consumers relinquish their right to reasonable privacy when they engage a third party such as Amazon and Alexa [18]. Privacy concerns will continue into the foreseeable future, especially with reports Amazon may release Alexa audio recordings to third-party developers [19].

Analysis of the two Echo Dot datasets, using the AVS signature, identified inconsistencies in the number of logged response cards to the number of AVS data points. In the 2017 dataset, three AVS data points were identified but only 2 original response cards were logged. Again, in the 2018 dataset, four AVS data points were identified but only three response cards were logged. On March 14, 2017, and April 22, 2018, ED1 logged one original response card but no associated AVS data point was identified in the network traffic. The identification of these AVS traffic and logged response card inconsistencies demonstrated that the Echo Dot did not properly log user commands for review by the consumer; therefore, research question 3 "Are Alexa streamed user commands properly logged and available for review by the consumer?" is answered in the negative.

One of the apparent limitations of this study was the robustness and reliability of the AVS signature. Even though the AVS signature was evaluated against two additional datasets, since k-mean cluster analysis is a descriptive methodology, it is unlikely it would retain its reliability when used with larger datasets. Even though the AVS signature identified Google Home audio data points, four data points is not sufficient to generalize about the reliability of the AVS 7050 bpm signature with non-Amazon devices. The unreliability of the Alexa application when reviewing response cards was another limitation of this study. Transcribing response cards daily and recording the response card audio to a different medium for later replay would have improved the response card documentation process. Another limitation of this study was knowledge gaps in AVS and Echo day to day operations. Assumptions were made about duplicate response card lack of network traffic, the identification of keep-alive traffic, and the purpose of Cluster 2 network traffic.

This study highlighted several user privacy, consent, and data control issues. The spurious or inadvertent audio streams are potential violations of privacy since the user does not provide consent by invoking the wake word. Also, it is unknown if any user data is passed to AVS via encrypted data to other Amazon domains. And finally, the lack of control and reliability (disappearing and reappearing) of response cards contribute to consumers' doubt in ubiquitous computing.

## 6 Conclusion

This study contributed to the Echo and Alexa network behavior knowledge domain by answering the three research questions and identifying inconsistencies in the number of logged response cards to AVS data points. An outcome of this study is the creation of more questions regarding Echo and AVS network traffic. Additional research in this area is necessary to determine if Echo voice training reduces the number of inadvertent audio recordings, do other Alexa services, such as ordering items from Amazon's store, generate a detectable network traffic pattern. What is the relationship between pindorma AVS and other Amazon domain encrypted traffic?

Do microphone settings affect traffic to device-metrics-us.amazon.com and dp-gw-na.amazon.com?

Future research, using machine learning and larger datasets, may offer insight into the complex nature of ubiquitous computing. Future intelligent personal assistant and ubiquitous device research will provide consumers and corporations with knowledge and awareness to make appropriate security decisions.

**Acknowledgements** This research project did not receive specific or general grant funding from agencies in the public, commercial, or not-for-profit sectors.

## References

1. Alexa, say what?! Voice-enabled speaker usage to grow nearly 130% this year. (2017). https://www.emarketer.com/Article/Alexa-Say-What-Voice-Enabled-Speaker-Usage-Grow-Nearly-130-This-Year/1015812. Accessed October 2017
2. Mcgee M (2017) Sales, awareness of Amazon Echo skyrocket in Q4, analysts say. http://marketingland.com/sales-awareness-amazon-echo-skyrocket-q4-analysts-say-204298. Accessed October 2017
3. Levy N (2017) Amazon's smart assistant Alexa now tops 7,000 skills, a 7X increase in 7 months. https://www.geekwire.com/2017/amazons-alexa-now-tops-7000-skills-a-7x-increase-in-7-months/. Accessed October 2017
4. Hetzner C (2017) VW joins Ford, BMW to offer Amazon's 'Alexa' voice service. http://www.autonews.com/article/20170120/copy01/301209965/vw-joins-ford-bmw-to-offer-amazons-alexa-voice-service. Accessed October 2017
5. Amazon Web Services Incorporated (2017) Announcing Alexa for Business: using Amazon Alexa's voice enabled devices for workplaces. https://aws.amazon.com/blogs/aws/launch-announcing-alexa-for-business-using-amazon-alexas-voice-enabled-devices-for-workplaces/. Accessed December 4 2017
6. Kharpal A (2017) Amazon's voice assistant Alexa could be a $10 billion 'mega-hit' by 2020: Research. https://www.cnbc.com/2017/03/10/amazon-alexa-voice-assistan-could-be-a-10-billion-mega-hit-by-2020-research.html. Accessed October 2017
7. Vanian J (2017) This is Amazon's latest effort to get Alexa into more devices. FortuneMagazine. http://fortune.com/2017/08/17/amazon-alexa-devices-software-kit/. Accessed November 2017
8. Luger E, Rodden T (2013) An informed view on consent for UbiComp. Proceedings of the 2013 ACM international joint conference on pervasive and ubiquitous computing:529–538
9. Hong JI, Landay JA (2004) Proceedings of the 2nd international conference on mobile systems, applications, and services. New York, NY, USA
10. Clinton I, Cook L, Banik S. A survey of various methods for analyzing the Amazon Echo https://vanderpot.com/Clinton_Cook_Paper.pdf. Accessed September 14 2017
11. Chung H, Park J, Lee S (2017) Digital forensic approaches for Amazon Alexa ecosystem. Digit Investig 22:S15–S25
12. Padilla M (2017) The Amazon Echo Dot has a dump-ramdump.sh script. vol 2017
13. Micaksica (2017) Exploring the Amazon Echo Dot, part 1: intercepting firmware updates. Medium
14. Apthorpe N, Reisman D, Feamster N (2017) A smart home is no castle: privacy vulnerabilities of encrypted IoT traffic. arXiv preprint arXiv:170506805
15. Beran B (2016) Understanding clustering in Tableau 10. https://boraberan.wordpress.com/2016/07/19/understanding-clustering-in-tableau-10/. Accessed April 24 2018
16. Tableau Support Page How clustering works. https://onlinehelp.tableau.com/current/pro/desktop/en-us/clustering.html#HowItWorks. Accessed April 26 2018
17. na (2014) K3 WiFi with no Internet access. http://www.mobileread.mobi/forums/showthread.php?t=237799
18. Edwards HS (2017) Alexa takes the stand: listening devices raise privacy issues. Time Magazine. http://time.com/4766611/alexa-takes-the-stand-listening-devices-raise-privacy-issues/. Accessed September 16 2017
19. Statt N (2017) Amazon may give app developers access to Alexa audio recordings. https://www.theverge.com/2017/7/12/15960596/amazon-alexa-echo-speaker-audio-recordings-developers-data. Accessed January 7 2018