University at Albany, State University of New York
Scholars Archive

Business/Business Administration | Honors College

Spring 5-2019

Smart Speaker Forensics

Steven Engelhardt
University at Albany, State University of New York, sengelhardt@albany.edu

Smart Speaker Forensics

An honors thesis presented to the
Department of Information Security and Digital Forensics,
University at Albany, State University of New York
in fulfillment of the requirements
for graduation from The Honors College.

Steven Engelhardt

Research Advisor: Fabio Auffant II

February 2018

Abstract

Devices such as the Google Home and Amazon Echo are great sources of digital evidence and artifacts due to the fact that much of the data these devices generate are stored both locally and in the cloud. For this reason, law enforcement and forensic investigators are not forced to rely on companies such as Google and Amazon to retrieve useful information. In this paper, techniques used to retrieve this information are discussed. Although some of these techniques require an analyst to have an individual's login credentials, other techniques that focus on the data stored in a user's phone can be used as long as the analyst is able to obtain root (privileged) access.

Acknowledgements

Many people have helped me through the process of writing my thesis. My thesis advisor, Fabio Auffant II was particularly helpful in reviewing my thesis and guiding me throughout the process of conducting my research. It was in his class that I first came up with the idea for my thesis.

I’d also like to thank my parents and my siblings for encouraging me to succeed and instilling in me the importance of a good education. Their help and guidance for the past 20 years has made me who I am today and none of this would have possible without them.

Table of Contents

Introduction

Internet of Things (IoT) devices are becoming commonplace in homes, schools and places of work. Individuals use IoT devices to make their lives easier through automation and the use of virtual assistants such as the Google Assistant and Amazon Alexa. Two devices in particular, the Google Home and Amazon Echo are increasingly popular due to their relatively low cost and the free cloud-based services associated with them.

The ubiquitous use of the Google Home and Amazon Echo has generated a large amount of data that investigators can use as a source of evidence. In the past several years, law enforcement officials and forensic analysts have attempted to use these devices to gather evidence but received pushback from device manufacturers. In 2015 for instance, James Bates was charged with first-degree murder. Investigators seized an Amazon Echo smart speaker from his home and requested that Amazon hand over "pertinent information regarding the device's communication with Alexa." Amazon denied law enforcements request, citing an absence of "valid and binding legal demand" (Chung, Park, & Lee, Digital Forensic Approaches for Amazon Alexa Ecosystem, 2017). Although Amazon's respect for user privacy is amiable, Amazon's behavior in this case and cases like it represent a challenge for investigators that try to retrieve evidence from IoT devices. This paper presents several approaches to examining and collecting evidence from the Google Home and Amazon Echo so that forensic analysts can rely less on device manufacturers and more on methods of retrieving data themselves.

Background

The Amazon Echo was first released in November 2014 (Lorenzetti, 2014). At the time this paper was written, a limited amount of research has been conducted on retrieving evidence from the Amazon Echo. Unfortunately, because new iterations of the Amazon Echo and its companion application are consistently released, some of this research discusses outdated methods and information.

The Google Home was first released in October 2016 (Kovach, 2016). Because the device is a much newer device than the Amazon Echo, very little research has been conducted on retrieving evidence from the Google home. In fact, at the time this paper was written, I was unable to find any prior digital forensic research conducted on the Google Home.

Purpose and Scope

This research project's primary goal was to learn everything I could about the Google Home and Amazon Echo. Because the two smart speakers function primarily through software, the project focused on examining their associated Android companion applications and cloud-based services (The Senator Patrick Leahy Center for Digital Investigation, 2016). The smart speakers' hardware was largely ignored.

Since research already existed on the Amazon Echo at the time the project started, I focused on verifying the methods used by other researchers to obtain their results and determining if any other forensic artifacts could be recovered. Existing knowledge of the Google ecosystem and Google Assistant was used to examine the Google Home.

Research Questions

- What forensic artifacts can be recovered from the applications and services associated with the Google Home?
- What forensic artifacts can be recovered from the applications and services associated with the Amazon Echo?
- Which methods used to recover forensic artifacts from the applications and services associated with the Amazon Echo are outdated and/or irrelevant?
- What methods can be used to preserve the forensic artifacts associated with the Google Home and Amazon Echo?
- Are there any similarities between the structure and availability of data between the two smart speakers?

Google Home

Initial Setup

Before collecting data, a Samsung Galaxy S5 was connected to a Google Home Mini using the Google Home App. This process involved powering on the Google Home Mini, following the voice prompts to connect the device to the same Wi-Fi network as the Samsung Galaxy S5 and then linking a Google account to the speaker. Although it is possible to set up the

Google Home Mini using an iOS device, I did not attempt to do so because it was outside the scope of the project.

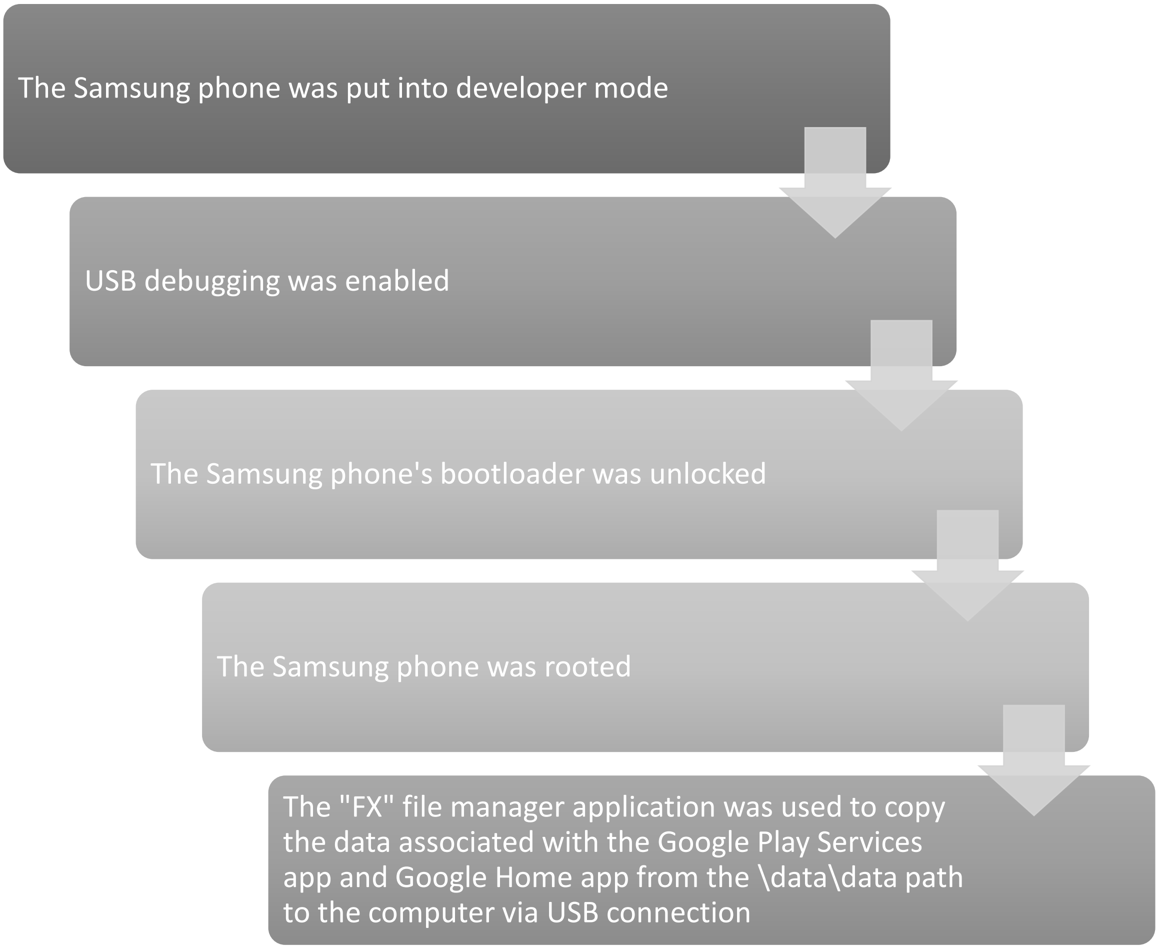
Equipment Used

Device	Operating System	Comments
Samsung Galaxy S5 (SM-G900V)	Android 6.0.1	Used to set up the Google Home Mini and collect data
Google Home Mini	Proprietary "Chromecast" based OS (Bohn, 2016)	
HP Spectre x360 Convertible 15-bl0XX	Windows 10 Home Edition (Version 1809)	Used to store copies of the collected data and analyze the digital artifacts from the companion applications

Software	Version	Comments
7-Zip	18.05 (x64)	
Android Studio	1.0	
DB Browser for SQL Lite	3.10.1	
Home App	1.30.43.17	com.google.android.apps.chromecast.app
Google Play Services App	14.3.66	com.google.android.gms
FX File Manager	7.2.2.2	nextapp.fx

Data Collection

After the Google Home application was installed on the Samsung phone, the Google Home was used for approximately 4 weeks. The Google Assistant, the voice-activated virtual assistant used to interact with the Google Home, was asked questions, asked to control music and other devices (such as TVs), set alarms, and set reminders among other things. At the conclusion of the 4-week period, 2 of the applications associated with the Google Home were extracted from the Samsung phone so that they could be analyzed. This extraction process has been outlined below:



The Samsung phone was put into developer mode

USB debugging was enabled

The Samsung phone's bootloader was unlocked

The Samsung phone was rooted

The "FX" file manager application was used to copy the data associated with the Google Play Services app and Google Home app from the \data\data path to the computer via USB connection

Analysis

After the application files were copied to the computer, an attempt was made to analyze each file using Android Studio, Notepad and/or a database viewer (DB Browser for SQL Lite).

Analysis of the Google Home App

When analyzing the Google Home application for potential digital artifacts, two files stood out. The first file, originally found in *data\data\com.google.android.apps.chromecast.app\shared_prefs\com.google.android.apps.chromecast.app_preferences_no_backup.xml*, has two strings that might be helpful to forensic investigators. One string, which was aptly named “current_account_name,” contained the email address of the user associated with the Google Home. In an investigation, this string might help investigators by providing them with the email of the Google Home’s primary user. The other string contains “current_home_id,” indicating that it is possibly a unique identifier for the

Google Home device. If this string is indeed an identifier, investigators might be able to use it to correlate the activity on a Google Home with a specific user.

The second file, originally found in *data\data\com.google.android.apps.chromecast.app\shared_prefs\com.google.android.apps.chromecast.app_preferences.xml*, contained 3 Boolean values and indicated whether or not the Google Home and Google Assistant were used previously. In the actual XML file examined, all 3 Boolean values were set to false, indicating that Google Home and Google Assistant were used at some point.

Analysis of the Google Play Services App

When analyzing the Google Play Services application for potential digital artifacts, an unencrypted database containing information about every reminder created with the Google Home was examined. This database, originally found in *data\data\com.google.android.gms\databases\Reminders.db,* has a table called "reminders" that indicates:

- Whether or not a reminder is time or location based
- The text transcription of the reminder
- The time/date the reminder was created
- The time/date the reminder was archived
- Whether or not the reminder was "deleted"
- The time/date that Google should notify the user about the reminder (reminder completion date)
- How often the reminder should occur

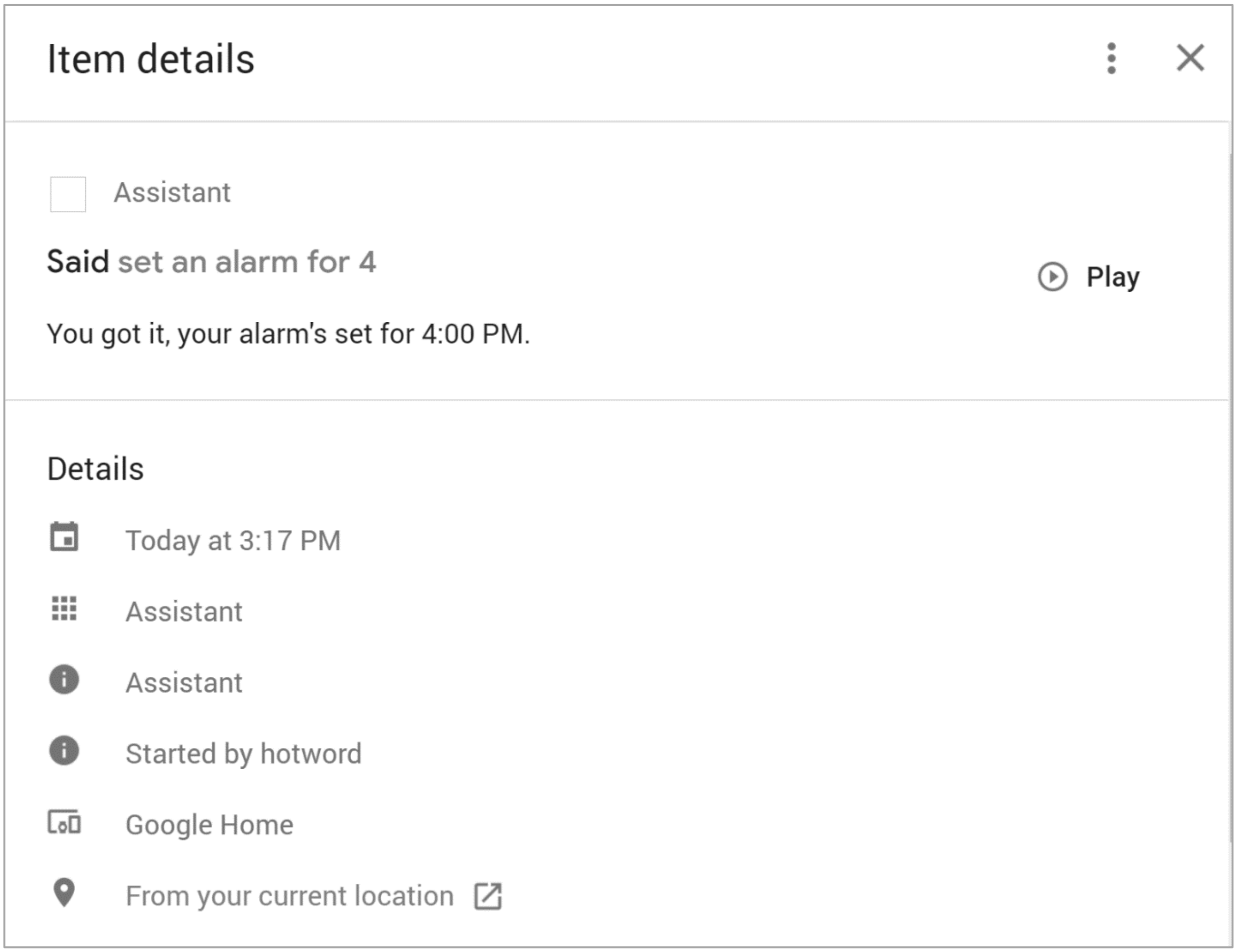
For reminders that are location based, the table has keys providing the location's exact coordinates and/or description. This information, combined with other information about the reminder, can be of practical use to investigators. For example, if a user were to create a location-based reminder using the Google Home, and arrive to the location specified using one of their Google devices, the reminder will likely be marked as completed. An investigator can use the record in the reminder.db database to help prove that a user arrived at the location at the time the reminder was marked as completed.

After clearing the application's data and reexamining the database, it was found that the database is not entirely cached or temporary. In other words, reinstalling the application results in the reminders being downloaded from Google’s servers again. As a result, a copy of all reminders is stored locally on Android phones allowing forensic examiners to access a historical record of all reminders without contacting Google.

Google Cloud Considerations

Since the Google Home is primarily a cloud-based service, a lot of digital artifacts are stored in the cloud and not locally. Accessing myactivity.google.com/myactivity with the Google account associated with a Google Home allows an individual to obtain some of these artifacts. Using this web page, it is possible to view text transcriptions of the commands given to a Google Home and play back audio recordings of said commands. In addition, it is possible to determine when a specific command was stated and the location of a user at the time the command was executed. Assuming that a forensic investigator has access to a user’s Google account, he or she can use this information to piece together a timeline of a user’s activity.

As noted by Ariel Watson, the audio recordings accessible on this webpage can not only be used by law enforcement to hear a user’s voice, but possibly unlock a user’s device if the “Okay Google” voice unlock feature is enabled (Watson, 2018). This feature allows an individual to unlock their device by stating “Okay Google.” Assuming one of the voice recordings from the My Activity page is clear enough, law enforcement can play back an audio recording and a device with the feature enabled should unlock itself. For time sensitive investigations or investigations where a device’s passcode is unknown, this method could prove incredibly useful.



Assistant

Said set an alarm for 4

You got it, your alarm's set for 4:00 PM.

Details

Today at 3:17 PM

Assistant

Assistant

Started by hotword

Google Home

From your current location

Figure 1 Sample of the myactivity.google.com/myactivity web page

At 3:17 PM on November 11th, 2018, the Google Home was asked to set an alarm for 4:00 PM. This is reflected in the text transcription and detailed timestamp. By clicking on the "Play" button, it is possible to play back an audio recording of this command. By clicking on "From your current location," it's possible to view the location that the command was initiated from on Google Maps.

Amazon Echo

Initial Setup

Before collecting data, a Samsung Galaxy S5 was connected to a second-generation Amazon Echo Dot using the Amazon Alexa App. This process involved powering on the Amazon Echo Dot, following the voice prompts to connect the device to the same Wi-Fi network as the Samsung Galaxy S5 and then linking an Amazon account to the speaker.

Equipment Used

Device	Operating System	Comments
Samsung Galaxy S5 (SM-G900V)	Android 6.0.1	Used to set up the Google Home Mini and collect data
Second-Generation Amazon Echo Dot	Fire OS	
HP Spectre x360 Convertible 15-bl0XX	Windows 10 Home Edition (Version 1809)	Used to store copies of the collected data and analyze the digital artifacts from the companion application

Software	Version	Comments
7-Zip	18.05 (x64)	
Android Studio	1.0	
DB Browser for SQL Lite	3.10.1	
Amazon Alexa App		com.google.android.apps.chromecast.app
FX File Manager	7.2.2.2	nextapp.fx

Data Collection

After the Amazon Alexa application was installed on the Samsung phone, the Echo Dot was used for approximately 2 weeks. Amazon Alexa, the voice-activated virtual assistant used to interact with the Echo Dot, was asked questions, asked to set reminders and set timers. At the conclusion of the 2-week period, the Amazon Alexa application associated with the Echo Dot was extracted from the Samsung phone so that they could be analyzed. With the phone already being rooted and USB debugging enabled, this extraction process involved copying the data associated with the Amazon Alexa app from the \data\data path to the computer via USB connection.

Analysis

Similar to the Google Home analysis, after the application files were copied to the computer, an attempt was made to analyze each file using Android Studio, Notepad and/or a database viewer (DB Browser for SQL Lite).

Analysis of the Amazon Alexa App

In the research paper “Digital Forensic Approaches for Amazon Alexa Ecosystem,” the authors discuss several files that contain client centric artifacts from the Amazon Alexa companion application (Chung, Park, & Lee, Digital Forensic Approaches for Amazon Alexa Ecosystem, 2017). One of these files, a database called DataStore.db (/data/data/com.amazon.dee.app/databases/DataStore.db), contained unparsed information related to the to-do and shopping lists created with the Echo. In my analysis of the Amazon Echo companion application, I parsed the same database and determined that the same information is still available with the current version of the application and Echo. Specifically, I found that the database contains a table called “DataItem” providing:

- A text transcription of the item/s added to a user’s Amazon Alexa shopping list
- A text transcription of the thing/s added to a user’s Amazon Alexa to-do list
- The date and time that an item was added to a user’s Amazon Alexa shopping list
- The date and time that something was added to a user’s Amazon Alexa to-do list
- The customer ID of the individual associated with the shopping and to-do lists

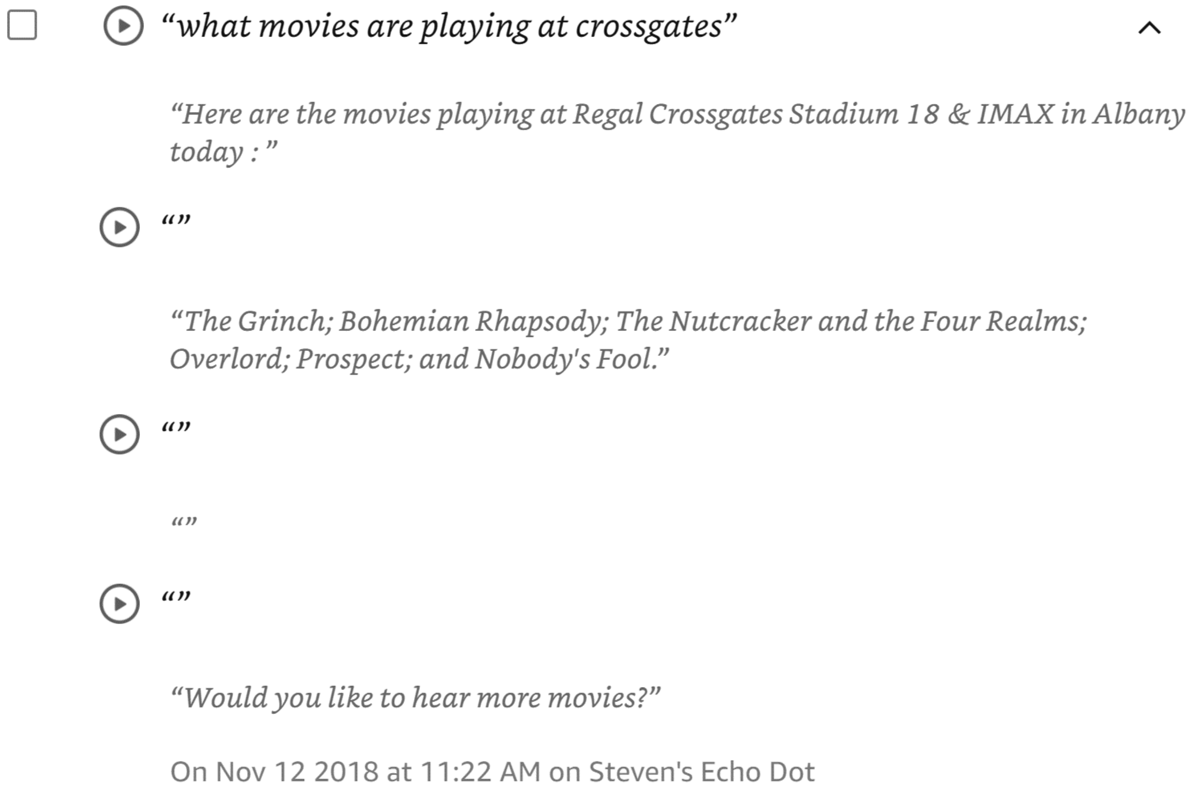
An example of a database cell from this database is included below, with key information useful to forensic investigators highlighted.

[{"completed":false,"createdDateTime":1542050207335,"customerId":null,"id":"**0a262498-b097-4133-a102-16bca2ed08fc**","listId":"YW16bjEuYWNjb3VudC5BSFJQU0haQjMyUERSNkRJWkRaTkRMTTRKWE5RLVNIT1BQSU5HX0lURU0=","shoppingListItem":true,"updatedDateTime":1542050207335,"value":"**crayons**","version":1},{"completed":false,"createdDateTime":**1542050163851**,"customerId":null,"id":"**f717a9cc-f540-4f35-b73b-2e15daa2f28b**","listId":"YW16bjEuYWNjb3VudC5BSFJQU0haQjMyUERSNkRJWkRaTkRMTTRKWE5RLVNIT1BQSU5HX0lURU0=","shoppingListItem":true,"updatedDateTime":1542050163851,"value":"**computer paper**","version":1}]

The same research paper indicated that the Amazon Echo companion application had a database called map_data_storage.db (Chung, Park, & Lee, Digital Forensic Approaches for Amazon Alexa Ecosystem, 2017). In my analysis of the application files, I determined the current version of the companion application had a similar database but called map_data_storage_v2.db. This database includes tokens associated with the active user, the number of accounts associated with the Echo, as well as the “display name” of the active user. In the database examined, this “display name” was my full name making it relatively easy for a forensic investigator to determine ownership of the device.

Amazon Cloud Considerations

As is the case with the Google Home, the Amazon Echo is primarily a cloud-based service. As a result, most of the digital artifacts associated with the device’s usage is stored in the cloud. Accessing amazon.com/alexaprivacy with the Amazon account associated with an Amazon Echo allows an individual to obtain some of these artifacts. Using this web page, it is possible to view text transcriptions of the commands given to an Amazon Echo and play back audio recordings of said commands. In addition, it is possible to determine when a specific command was stated. Unlike the with Google “My Activity” page however, it is not currently possible to determine where the command was stated.



□ ▶ *“what movies are playing at crossgates”* ^

“Here are the movies playing at Regal Crossgates Stadium 18 & IMAX in Albany today : ”

▶ *“”*

“The Grinch; Bohemian Rhapsody; The Nutcracker and the Four Realms; Overlord; Prospect; and Nobody's Fool.”

▶ *“”*

“”

▶ *“”*

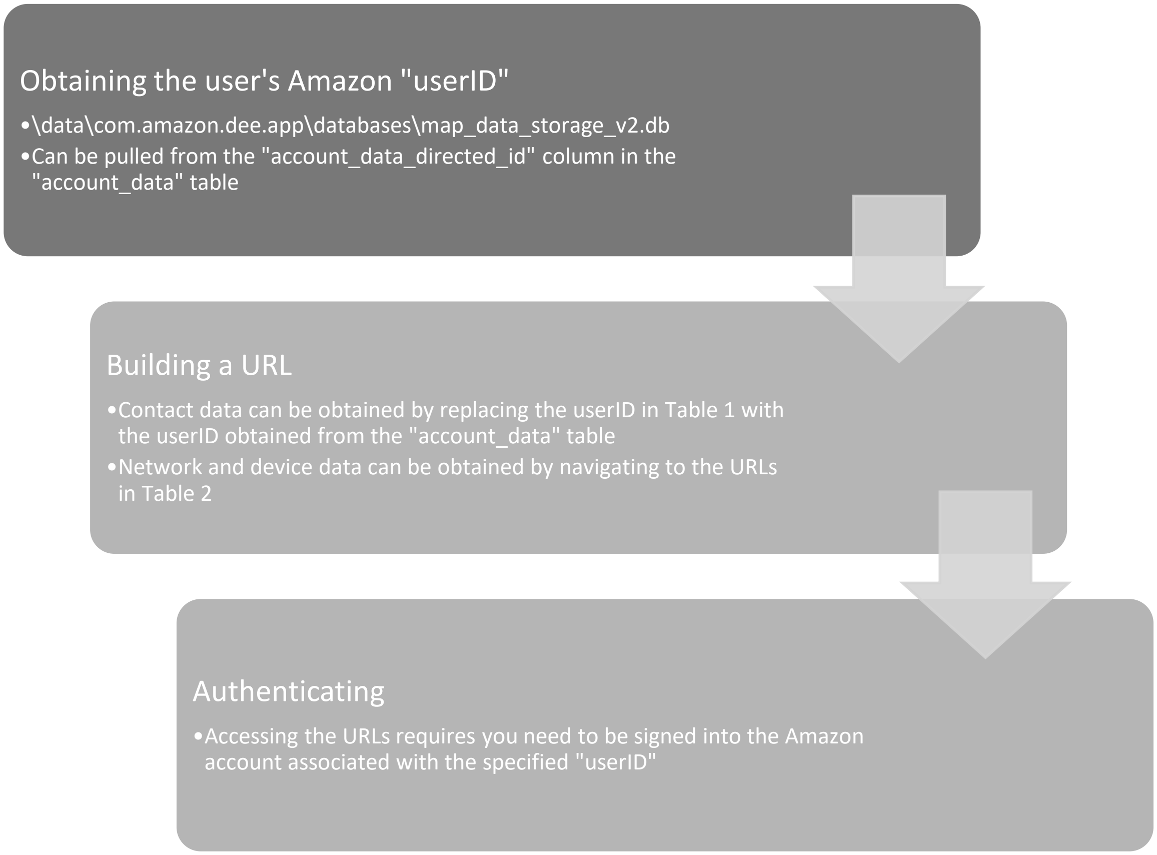
“Would you like to hear more movies?”

On Nov 12 2018 at 11:22 AM on Steven's Echo Dot

Figure 2 Sample of the amazon.com/alexaprivacy web page.

This page lists a detailed list of commands/questions given to the Amazon Echo and transcribes them.

As noted by Jessica Hyde in Magnet Forensics' "Alexa Cloud Data Reference Guide," it is also possible to retrieve sensitive contact, device and network information from Amazon's cloud (Moran, 2017). This can be done using the three-step process outlined below:



Obtaining the user's Amazon "userID"

- \data\com.amazon.dee.app\databases\map_data_storage_v2.db
- Can be pulled from the "account_data_directed_id" column in the "account_data" table

Building a URL

- Contact data can be obtained by replacing the userID in Table 1 with the userID obtained from the "account_data" table
- Network and device data can be obtained by navigating to the URLs in Table 2

Authenticating

- Accessing the URLs requires you need to be signed into the Amazon account associated with the specified "userID"

Conclusion

Upon completion of my research, I was able to obtain varying amounts of user data from the applications associated with the Google Home and Amazon Echo. Forensic analysts can use this data and the information in this report to aid in their investigations.

Due to the frequent software and hardware updates that the Google Home and Amazon Echo receive, the information and user data retrieved during this project may be different several weeks or months from now. For this reason, it's important for forensic analysts to verify the methods outlined in this paper in the future to verify that they are still applicable to current iterations of the devices. Investigating the devices in the future might also yield additional results.

Future Work

The analysis of the Google Home and Amazon Echo in this project revolved primarily around software. Future research on Google Home and Amazon Echo forensics investigating the devices’ hardware could prove useful to investigators.

Most current forensic research into the Google Home and Amazon Echo collect data primarily from Android devices. It would be interesting to conduct more forensic research using applications installed on an iOS device, where file paths, application structure and data is likely to be different.

Appendix

Table 1

Data URLs (Moran, 2017)		**Comments**
Contacts	https://alexa-mobile-service-na-preview.amazon.com/users/amzn1.comms.id.person.amzn1~**amzn1.account.AF5YMXCV6JDBDMEYSGY67GYJWTBD**/contacts	Contains the user's phone number, full name, synced contacts, their names and their phone numbers.

Table 2

Data URLs (Moran, 2017)		**Comments**
Devices	https://pitangui.amazon.com/api/devices/devices?	Contains unique identifiers for the Amazon Echo, information about its owner, enabled services and software information.
Network	https://pitangui.amazon.com/api/wifi/configs?	Information about the network the Amazon Echo is connected to. Contains the network's security method, SSID and password.

Glossary

Acquisition – The process of copying data from a piece of evidence, to another location in a forensically sound manner so that the data may be analyzed at a later time. The goal is to leave the original media intact while working on a copy of it. This allows for evidence to be verified at a later date (The Senator Patrick Leahy Center for Digital Investigation, 2016).

Android (OS) – An open source operating system developed by Google and based on the Linux kernel for mobile devices with support for an expanding number of hardware devices (The Senator Patrick Leahy Center for Digital Investigation, 2016).

Android Studio – The official integrated development environment for Android platform development. It allows data to be pushed and pulled from an Android Device (The Senator Patrick Leahy Center for Digital Investigation, 2016).

APK – The android application format (The Senator Patrick Leahy Center for Digital Investigation, 2016).

Artifacts – Any data generated by user interaction that can be collected and examined. Any user data retrieved from the browser is considered an artifact, including cookies, caches, geolocation, search history, etc. (The Senator Patrick Leahy Center for Digital Investigation, 2016).

Parse – The process of dividing a computer language statement into parts that can be made useful for the computer. A parser in a program compiler is a program that takes each program statement that a developer has written and divides it into parts (for example, the main command, options, target objects, their attributes, and so forth) that can then be used for developing further actions or for creating the instructions that form an executable program (The Senator Patrick Leahy Center for Digital Investigation, 2016).

Rooting- Rooting enables a normal user to have administrator-level permissions to the operating system environment. In the case of Android devices, it helps in circumventing the security architecture (Techopedia, 2018).

7-zip – An open source software used primarily to compress and extract files. It has a command line and graphic user interface (The Senator Patrick Leahy Center for Digital Investigation, 2016).

References

Bohn, D. (2016, May 21). No surprise, Google Home is based on Chromecast, not Android. *The Verge.* Retrieved from https://www.theverge.com/circuitbreaker/2016/5/31/11822032/google-home-chromecast-android

Chung, H., Park, J., & Lee, S. (2017). Digital forensic approaches for Amazon Alexa ecosystem. *Digital Investigation*, *22*. doi:10.1016/j.diin.2017.06.010

Kovach, S. (2016, October 4). *Google unveils its newest major product: the Google Home speaker*. Retrieved from https://www.businessinsider.com/google-home-announced-price-release-date-2016-10

Lorenzetti, L. (2014, November 6). *Forget Siri, Amazon now brings you Alexa*. Retrieved from http://fortune.com/2014/11/06/forget-siri-amazon-now-brings-you-alexa/

Moran, B. (2017, December 26). *Alexa cloud data reference guide.* Retrieved from https://www.brimorlabsblog.com/2017/12/amazon-alexa-forensic-walkthrough-guide.html

Rooting. (n.d.). Retrieved from https://www.techopedia.com/definition/31284/rooting-smartphones

The Senator Patrick Leahy Center for Digital Investigation. (2016). *Amazon Echo Forensics.* Burlington, VA: Champlain College.

Watson, A. (2018, September 4). *Ok Google is more than OK for digital forensics investigations*. Retrieved from https://www.cellebrite.com/en/blog/ok-google-is-more-than-ok-for-digital-forensics-investigations