Alexa, what did I do last summer?

Vladimir Katalov, ElcomSoft

SecTor 2018
Amazon Alexa is a virtual assistant developed by Amazon.

- She’s 4 years young.
- First appeared in Amazon Echo in Nov 2014.
- Developed by Amazon Lab126.
Alexa

What Can She Do?

- Alexa is the smartest AI on the market
- She can talk!
  - Context-aware support added late 2018
- She can listen!
- She has skills!
- Play music, radio and podcasts, recommend playlists
- Read audio books

*Competitors: Apple Siri, Google Home, Microsoft Cortana*
Alexa

What Can She Do?

- Smart Home control and Smart Home automation
  - The largest number of compatible devices
- News, traffic, weather
  - Location aware!
- Make and receive calls
  - Alexa-to-Alexa, Drop-in, video calls
- She can sell you stuff!
  - Amazon is the largest online retailer
  - Have you seen anyone buying with Alexa?
Alexa

Alexa Is Always Listening

- Unless mic is manually disabled, she’s always listening
- Unless camera disabled, she’s always looking (Echo Spot, Echo Show)
- Does it pose a risk?
  - Yes: Alexa leaked private conversations, called random contacts
Alexa

Alexa Is Always Recording?

- No.
- Alexa waits for wake word
- If wake word is detected, she starts recording
- Recording sent to the cloud for voice recognition
- Both the voice sample and recognized command are saved in the user’s Amazon Account
Alexa

Devices and Alexa

- Amazon has device and software infrastructure
- **FireOS tablets** with Alexa: Kindle Fire, Fire HD, Fire HDX
- **FireTV** with Alexa
- Multiple generations of **Echo** smart speakers: Amazon Echo, Echo Plus, Echo Dot, Echo Show, Echo Sub, Echo Spot and more
Alexa

“Alexa, tell me about your new devices”

- Amazon Fire Phone
- Alexa Microwave
- Alexa Smart Plug
- Alexa on third-party devices (Sonos One, Moto Mods etc.)
- Alexa app (Android, iOS)
Alexa

How Does It Work?

▪ User speaks activation word
  ("Alexa", "Computer" or "Amazon")
▪ Alexa starts recording
▪ User speaks a command to Alexa-compatible app or device
▪ Device or app sends the audio recording to Amazon servers
▪ Sound recordings are decoded and processed with speech recognition
▪ Keywords and phrases are detected
1. User asks Alexa to do something

2. Device sends voice to Amazon cloud

3. Amazon server decodes voice and matches it with utterances

4. Amazon server sends JSON info via HTTPS to Amazon-compatibility service

5. JSON info with status and text of response

6. Device speaks response

Amazon Echo device or user’s Android/iOS-device with installed Amazon Alexa

Amazon-compatibility service such as: Amazon Music, Spotify, TuneIn, iHeartRadio or Smart Home System
Alexa

The Back End

- Amazon server prepares JSON with command code
- The command is sent to the relevant skill handler or service
  - Amazon Music or Spotify, smart home system, Alexa skill, IEEE routine etc.
- Service or skill handler attempts to execute the command
- Reports to Amazon server with status and text to be voiced back to the user
- Amazon server uses text-to-speech engine to make voice response
- The voice recording is sent to the user's device or app
- The device speaks out the answer from Amazon Alexa
What Does Alexa Know About You?

- Account & device settings
- Amazon order history, timeline, wish list
- Billing & shipping addresses, payment methods
- Installed applications, subscriptions, services
- Games, video library
- Voice snippets (mp3 + recognized commands)
- Voice calls (some countries), messages
Alexa

Amazon Hardware: Cloud Backups

- Amazon tablets and Fire Phone feature cloud backups
- Even if Android version (FireOS) lower than 6.0
- Separate from Google infrastructure
- List of installed apps
- Photos and documents
- Wi-Fi passwords
- Silk browser data
Alexa

Where Do You Keep My Data?

- In AWS cloud
- Amazon Alexa gets smarter with every request
- Alexa is always learning
- Alexa is constantly adapting to the user
- Developing recognition, analysis and responses requires real data
- User’s voice recordings and recognized requests are stored on Amazon servers
Alexa

Where Is My Data?

- Depends on device
- Echo and third-party Alexa speakers:
  - AWS
- Alexa app on Amazon and non-Amazon hardware:
  - AWS
  - Private app storage (sandboxed)
Alexa

Extracting Data from Alexa App

- iOS:
  - local backups (iTunes)
  - iCloud backups
  - /library/caches/com.amazon.echo/

- Android:
  - Need root access
  - Main database: com.amazon.echo
  - Additional tokens and authentication data; user ID for Mobile Service API:
    com.amazon.dee.app\databases\map_data_storage.db
# Alexa App Files

<table>
<thead>
<tr>
<th>OS</th>
<th>Application</th>
<th>Path</th>
<th>Format</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Android 4.4.2</td>
<td>Alexa 1.24.1176.0</td>
<td>/data/data/com.amazon.dce.app/databases/map_data_storage.db</td>
<td>SQLite</td>
<td>Tokens of an active user</td>
</tr>
<tr>
<td></td>
<td></td>
<td>/data/data/com.amazon.dce.app/databases/DataStore.db</td>
<td>SQLite</td>
<td>Todo and shopping list</td>
</tr>
<tr>
<td></td>
<td></td>
<td>/data/data/com.amazon.dce.app/app_webview/Cache/*</td>
<td>WebView cache</td>
<td>Cached native artifacts</td>
</tr>
<tr>
<td>iOS 10.1.1</td>
<td>Alexa 1.24.1176.0</td>
<td>[iTunes backup]/com.amazon.echo/Documents/LocalData.sqlite</td>
<td>SQLite</td>
<td>Todo and shopping list</td>
</tr>
<tr>
<td>OS X 10.10.5</td>
<td>Chrome 55.0.2883.87</td>
<td>~/Library/Caches/Google/Chrome/Default/Cache/</td>
<td>Chrome cache</td>
<td>Cached native artifacts</td>
</tr>
</tbody>
</table>
Alexa

**ZDATAITEM Documents\LocalData.sqlite**

- Data in JSON format (iOS/Android app)
  - Shopping items ordered via Alexa app
Alexa

Alexa and iOS Keychain

- Authentication data stored in the keychain
- Item: com.amazon.echo
- How to obtain iOS keychain:
  - Local backup (password-protected)
  - iCloud Keychain (Apple ID/password, 2FA, passcode or system password of an enrolled device)
  - Physical acquisition (jailbreak required!)
Amazon Alexa Authentication

- Web approach (as in iOS/Android apps)
- FIRS login
- Token-based login
Web Authentication

- Web authentication is used by iOS and Android apps
- Uses OpenId2.0 protocol, REST architecture
- Some information received via cookies
- Some data must be parsed off HTML pages

FIRS Host Alexa Authentication

- Requests sent to firs-ta-g7g.amazon.com/FirsProxy
- Implemented in a number of open source tools (lolsborn/readsyn)

Token-based

- User devices contain multiple authentication tokens
- Login/password authentication requests employ cookies with very short lifespan
- There might be a way to use tokens for generating such cookies
Alexa

Using Alexa API to Access Information

- Most information is available with login and password (and potentially 2FA)
  - Requests history, recorded voice requests, translated requests
  - List of devices registered in the account
  - Payment cards
  - Wi-Fi data
  - Smart home devices
  - Named lists
  - Household data
  - Amazon Music data
  - Wake word
- https://alexa.amazon.com/api
Using Alexa API to Access Information

<table>
<thead>
<tr>
<th>Activities</th>
<th><a href="https://alexa.amazon.com/api/activities">https://alexa.amazon.com/api/activities</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>WAV-records</td>
<td><a href="https://alexa.amazon.com/api/utterance/audio/data?id=">https://alexa.amazon.com/api/utterance/audio/data?id=</a></td>
</tr>
<tr>
<td>Devices</td>
<td><a href="https://alexa.amazon.com/api/devices/device">https://alexa.amazon.com/api/devices/device</a></td>
</tr>
<tr>
<td>WiFi</td>
<td><a href="https://alexa.amazon.com/api/wifi/configs">https://alexa.amazon.com/api/wifi/configs</a></td>
</tr>
<tr>
<td>Smart home devices</td>
<td><a href="https://alexa.amazon.com/api/phoenix">https://alexa.amazon.com/api/phoenix</a></td>
</tr>
<tr>
<td>Named lists</td>
<td><a href="https://alexa.amazon.com/api/namedLists">https://alexa.amazon.com/api/namedLists</a></td>
</tr>
<tr>
<td>Householder info</td>
<td><a href="https://alexa.amazon.com/api/household">https://alexa.amazon.com/api/household</a></td>
</tr>
</tbody>
</table>
Alexa

Alexa Mobile Service API

- Mobile Service API is available in addition to Amazon REST API
- Available data: conversation logs with other Alexa/Echo users
- [https://alexa-mobile-service-na-preview.amazon.com/users/amzn1.comms.id.person.amzn1~](https://alexa-mobile-service-na-preview.amazon.com/users/amzn1.comms.id.person.amzn1~)
Using Alexa Mobile Service API

- Obtain User ID
  - From `map_data_storage_v2.db` (data\com.amazon.dee.app\databases)
  - From Amazon account (login/password)
- GET: `https://api.amazon.com/user/profile`

- User devices contain multiple authentication tokens
- Login/password authentication requests employ cookies with very short lifespan
- There might be a way to use tokens for generating such cookies

<table>
<thead>
<tr>
<th>Contacts</th>
<th><code>https://alexa-mobile-service-na-preview.amazon.com/users/amzn1.comms.id.person.amzn1-userId/contacts</code></th>
</tr>
</thead>
<tbody>
<tr>
<td>Conversations</td>
<td><code>https://alexa-mobile-service-na-preview.amazon.com/users/amzn1.comms.id.person.amzn1-userId/conversations</code></td>
</tr>
<tr>
<td>Dialog messages</td>
<td><code>https://alexa-mobile-service-na-preview.amazon.com/users/amzn1.comms.id.person.amzn1-userId/conversations/ConversationId/messages?count=</code></td>
</tr>
</tbody>
</table>
Now you can call and message your friends and family that have Echo devices. To set up additional members of your family, download and install the app on their phone.
Extracting Recorded Voice Snippets

- Accessing recorded voice snippets requires recording ID
- Recording ID is returned by requesting the list of activities
- Voice snippets are returned in base64-encoded WAV files
- All other requests return data in JSON format
- Description: parsed text of the voice request and voice snippet ID
- `activityStatus`:
  - SUCCESS
  - DISCARDED_NON_DEVICE_DIRECTED_INTENT (device didn’t recognize the request)
  - SYSTEM_ABANDONED (request interrupted)
- `utteranceld` – the ID to access the snippet
Amazon Drive: Authentication

- Amazon Drive uses RESTful API
- Authentication requires user authorization through Login with Amazon
- RESTful API allows authorized apps accessing photos, videos and documents stored in Amazon Drive
Alexa

Amazon Drive: Accessing the Data

- Information about the user
- Account info
- Data returned in JSON format
- Generate link request
- File links returned as a result
- Raw files can be downloaded using returned links
Proof of Concept: The Tool

- Requires Amazon login and password credentials
- Makes use of several different APIs to access different types of data
- Credentials are temporarily cached and reused to authenticate across various APIs
- Supports 2FA (wasn’t easy)
- 2FA codes are time bound (TOTP, 30 seconds)
- Several 2FA requests by different APIs
- Fresh 2FA codes requested when required
Alexa

Proof of Concept: The Tool

- **AmazonData**: authentication and data access routines implemented as a library
- **AmazonTestApp**: command-line tool that makes use of AmazonData to authenticate and download data
- **Web-based authentication via Login with Amazon**
- **Web API**:
  - Data in JSON format
  - Files (photos, videos, documents)
  - Voice command snippets in WAV format
- **Amazon Mobile API**:
  - Contacts
  - Conversations
  - Messages
Enter two factor code
019772
Logged in
Download alexa data...
Downloading Devices Data..done
Downloading Cards..done
Downloading Wifi Config..done
Downloading Smart Home data..done
Downloading Conversations..done
Downloading Contacts..done
Downloading Named Lists..done
Downloading Household..done
Downloading Music Data..done
Downloading Wake word..done
Downloading Skills..done
Downloading Device Preferences..done
Downloading Activities...done
Saving 22
..................done
all done
Downloading amazon drive data...
Downloading Account Info..done
Downloading Account Endpoint..done
Downloading Account Quota..done
Downloading Account Usage..done
stopped
Need another auth for www.amazon.de
Using your credentials to authorize
Enter two factor code
Alexa\Activities
Alexa\Cards.json
Alexa\Conversations
Alexa\Conversations.json
Alexa\DevicePreferences.json
Alexa\DevicesData.json
Alexa\Household.json
Alexa\list.txt
Alexa\MusicAccountDetails.json
Alexa\NamedLists.json
Alexa\Skills.json
Alexa\SmartHomeData.json
Alexa\WakeWord.json
Alexa\WiFiConfig.json
Alexa\Activities\A1DGPTNHQDWH11#1523996496415#A3NWHXTQ4EBCZS#AD7491D22EDD49A2ADF636E6EADD13E8
Alexa\Activities\A1DGPTNHQDWH11#1530918107538#A2IVLV5VM2W81#4F32176797174237820292ABB4382354
Alexa\Activities\A1DGPTNHQDWH11#15309181110304#A2IVLV5VM2W81#4F32176797174237820292ABB4382354
Alexa\Activities\A1DGPTNHQDWH11#1530918113350#A2IVLV5VM2W81#4F32176797174237820292ABB4382354
Alexa\Activities\A1DGPTNHQDWH11#1530918113650#A2IVLV5VM2W81#4F32176797174237820292ABB4382354
Alexa\Activities\A1DGPTNHQDWH11#1530918113650#A2IVLV5VM2W81#4F32176797174237820292ABB4382354\audio.wav
Alexa\Activities\A1DGPTNHQDWH11#1530918113650#A2IVLV5VM2W81#4F32176797174237820292ABB4382354\text.json
Alexa\Activities\A1DGPTNHQDWH11#1530918107538#A2IVLV5VM2W81#4F32176797174237820292ABB4382354\audio.wav
Alexa\Activities\A1DGPTNHQDWH11#1530918110304#A2IVLV5VM2W81#4F32176797174237820292ABB4382354\audio.wav
Alexa\Activities\A1DGPTNHQDWH11#1530918110304#A2IVLV5VM2W81#4F32176797174237820292ABB4382354\text.json
Alexa\Activities\A1DGPTNHQDWH11#1530918113650#A2IVLV5VM2W81#4F32176797174237820292ABB4382354\audio.wav
Alexa\Activities\A1DGPTNHQDWH11#1530918113650#A2IVLV5VM2W81#4F32176797174237820292ABB4382354\text.json
Alexa\Conversations\amzn1.comms.messaging.id.conversation~"-loFkQx4XiIkLoeDkAWSX9BVmcl.json
Alexa\Conversations\amzn1.comms.messaging.id.conversation~gp7yFMxi9gK26UbmyacbVF7GumA.json

Downloaded Files
<en-US>
    <utterances>
        &quot;Alexa, turn on the coffee maker;&quot;, &quot;Alexa, set bedroom light to green;&quot;, &quot;Alexa, make kitchen light cooler;&quot;)
    </utterances>
</en-US>

<fr-FR>
    <utterances>
        &quot;Alexa, allume le sapin de Noël;&quot;, &quot;Alexa, allume le salon en vert;&quot;, &quot;Alexa diminue la luminosité de la chambre;&quot;)
    </utterances>
</fr-FR>
<vendorId>MAZV0000</vendorId>
<consentUnavailable>true</consentUnavailable>
<lazyAccountLinking>false</lazyAccountLinking>
<skillTypes>SMART_HOME</skillTypes>
<supportsAccountLinking>true</supportsAccountLinking>
</productDetails>
<isInSkillProduct>false</isInSkillProduct>
<detailPageUrl>/gp/product/A2W9SC0A7O5T</detailPageUrl>
<productMetadata>
<skillId>amzn1.ask.skill.86c3.12s51.1.3.5@2019-04-15-08.15.20</skillId>
<badge/></badge>
<stage>live</stage>
<enablementPartnerId>QkhwUtA0QdGc</enablementPartnerId>
<stageLastUpdatedTime>152883138565</stageLastUpdatedTime>
<status>PUBLISHED</status>
</productMetadata>
<preRenderedIconUrl>https://s3.amazonaws.com/CAPS-
SSE/echo_developer/9949f9d4738a794e36/APP_ICON?versionId=p
</preRenderedIconUrl>
<hasInSkillProducts>false</hasInSkillProducts>
<hideReviews>false</hideReviews>
<utterancesByLocale>
<block data="" data-priority="default"/>
<contents>
  <id>tokenData</id>
  <contents>
    <csrfToken>AlcGf2G2L7d71G2SmM1AfCJi4</csrfToken>
  </contents>
  <contents>
    <id>kstData</id>
    <contents>
      <kstWidgetPath>#settings/kidSkills</kstWidgetPath>
      <kstStatus>false</kstStatus>
      <kstLearnMoreUrl>help/node/0FAAD3C5YV6A</kstLearnMoreUrl>
    </contents>
  </contents>
</contents>
<contents>
  <id>skillsPageData</id>
  <etag>1111e8d259b02d70629f7c3gKw</etag>
  <contents>
    <mode>flat</mode>
    <cards>
      <tabId>your-skills-tab-all</tabId>
      <formattedSkillCount>8</formattedSkillCount>
      <skillsCardPriority>neutral-priority</skillsCardPriority>
    </cards>
  </contents>
</contents>
Obtaining the Credentials

How to get cloud password or token?

- Legally (court order)
- Social engineering
- From computer (cached browser passwords)
- From computer (saved token from system or apps)
- Extract macOS keychain
- From other account that was easier to break (Apple / Google / Microsoft)
- Extract from local iTunes backup (with password)
- From password manager (need to crack master password first)
- Password re-use often helps
- From the sticker on monitor or note under the keyboard
- Rubberhose cryptanalysis
Alexa, what did I do last summer?

Vladimir Katalov, ElcomSoft