Digital Forensic Approaches for Amazon Alexa Ecosystem

DFRWS USA 2017

Hyunji Chung, Jungheum Park, Sangjin Lee
Korea University
## DFRC Research team

<table>
<thead>
<tr>
<th>Name</th>
<th>Position</th>
</tr>
</thead>
</table>
| **Hyunji Chung**| • Ph.D candidate in Korea University  
                  • Foreign researcher in National Institute of Standards and Technologies |
| **Jungheum Park**| • Ph.D in Korea University  
                         • Foreign researcher in National Institute of Standards and Technologies |
| **Sangjin Lee**  | • Professor in Korea University  
                         • Director of Digital Forensic Research Center |
Digital Forensics in IoT world

• IoT (Internet of Things) world
  • The network of physical objects that contain embedded communication technology
  • The worldwide IoT market: $1.7 trillion in 2020

• Digital Forensics in IoT world
  • Wearables, smart cameras, smart appliances
    → a large amount of digital data (great source of digital evidence)
  • Cloud based IoT devices

- Smart speaker
- Fitness tracker
- IP camera
- Connected car
Intelligent Virtual Assistant and Digital Forensics

- 25% of households using an intelligent virtual assistant (IVA) will have two or more devices by 2020
- Amazon Alexa-related environment will become an important source of potential digital evidence (CES 2017)

Real case related to **Amazon Echo** (Nov 2015)
- Police in Arkansas seized Bates’ Echo from his home
- asked Amazon to hand over any pertinent information regarding the device’s communication with Alexa cloud
- However, Amazon denied the request in the absence of a valid and binding legal demand
Amazon Alexa and Digital Forensics

• Amazon Alexa Ecosystem

- Alexa-enabled devices
  - Echo
  - Dot

- Alexa Companion Clients
  - Smart Devices
  - Laptop
  - PC

- Hardware
- Network

- Cloud

- Other Clouds
  - Smart Lighting Kits
  - Smart Smoke Alarm

- Compatible IoT Devices
  - Pizza Delivery
  - Ridesharing

- Third-party Applications

- User
- Network
Digital Forensic Analysis Strategy

Cloud-native artifacts using unofficial APIs

Web debugging proxy

Web browser

Mobile app

Hardware

eMMC Root
JTAG
Debug ports

Alexa-enabled devices
Echo
Dot

Alexa

Network

Network

Client

Smart Devices
Laptop
PC

Companion Clients

Digital Forensic Analysis Strategy
Related works

**IoT forensics**
- Hypothetical IoT crime scenarios (Oriwoh et al.)
- Definition of IoT forensics (Zawoad et al.)

**Cloud forensics**
- Fundamental challenges for IoT forensics (Hegarty et al.)
- Forensic framework in the IoT domain (Kebande et al.)
- Cloud-native forensics (Vassil et al.)
- Client-centric cloud forensics (Hyunji et al., Hale, Martini et al., Quick et al.)

**Digital forensics for IVA ecosystem**
- combine two perspectives on cloud forensics in order to propose an integrated IoT forensic system for the Amazon Alexa ecosystem
- cloud-native forensics is essential for identifying user behaviors
- client-centric forensics can enhance results of cloud-native forensics
Forensic Artifacts on Amazon Alexa Ecosystem (1/4)

• Test environment

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alexa-enabled devices</td>
<td>(1) Echo Dot (S/N: <em><strong>0L9</strong></em>473***P)</td>
</tr>
<tr>
<td></td>
<td>(2) Echo Dot (S/N: <strong>90</strong>*964*****U)</td>
</tr>
<tr>
<td></td>
<td>* some characters of S/N are masked by asterisks</td>
</tr>
<tr>
<td>Companion clients and applications</td>
<td>(1) Android 4.4.2 + Alexa app (1.24.1176.0)</td>
</tr>
<tr>
<td></td>
<td>(2) iOS 10.1.1 + Alexa app (1.24.1176.0)</td>
</tr>
<tr>
<td></td>
<td>(3) OS X 10.10.5 + Chrome (55.0.2883.87)</td>
</tr>
<tr>
<td></td>
<td>(4) Windows 10 + Chrome (55.0.2883.87)</td>
</tr>
<tr>
<td>Total test period</td>
<td>2016-11-18 ~ 2017-01-29</td>
</tr>
<tr>
<td>Last verification date</td>
<td>2017-08-02</td>
</tr>
</tbody>
</table>
Forensic Artifacts on Amazon Alexa Ecosystem (2/4)

- Cloud native artifacts

  Revealing unofficial Alexa APIs
  - Understanding the communication protocols
  - We performed an intensive traffic analysis
    - Most traffic is transferred over encrypted connections
  - Valid user credential → we can identify additional APIs and interpret return values from Alexa

Alexa’s native artifacts

- Seven categories of APIs
  - account, customer setting, Alexa-enable device, compatible device, skill, user activity, etc.
- Forensically meaningful native artifacts from Alexa
  - registered user accounts, saved Wi-Fi settings (including unencrypted passwords)
  - Alexa-enabled devices, linked Google calendars, installed skill lists, user activity, user's actual voice

- Understanding the communication protocols and data formats used by Alexa
  - We performed an intensive traffic analysis using a web proxy
  - Most traffic is transferred over encrypted connections, native artifacts are returned as JSON format

- Valid user credential → we can identify additional APIs and interpret return values from Alexa

Revealing unofficial Alexa APIs
### Forensic Artifacts on Amazon Alexa Ecosystem (3/4)

- **Client-centric artifacts – Databases of the Alexa mobile app (1/2)**

<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.dee.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.dee.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.dee.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>
<tr>
<td>Windows 10</td>
<td>Chrome 55.0.2883.87</td>
<td>%UserProfile%\AppData\Local\Google\Chrome\User Data\Default\Cache\</td>
<td>Chrome cache</td>
<td>Cached native artifacts</td>
</tr>
</tbody>
</table>
Forensic Artifacts on Amazon Alexa Ecosystem (4/4)

- Client-centric artifacts – Android WebView cache & Chrome web cache (2/2)

Android WebView cache

(a) WebView Cache Internals

(b) Alexa cache as an example
Design and Implementation

- Cloud-based IoT Forensic Toolkit (CIFT)
  - The event flow diagram

**INPUT 1**
CLOUD, ID, Password

**INPUT 2**
COMPANION_APP_ANDROID, Directory Path

**INPUT N**
COMPANION_BROWSER_CHROME, Directory Path

**Cloud Native Module**
- Acquire (download) cloud native artifacts
- Unofficial APIs
- URLs found by DPM

**Companion Client Module**
- Process Android App
- Process iOS App
- Process Chrome Cache

**Set up Environment**
Create a Web Session
Add a user-input
Process user-inputs

**Result DB**
- Acquired File
- Account, Device, Setting
- Timeline

**Evidence Library**
- JSON (Alexa API)
- SQLite (Alexa app)
- WAV (User's voice)

**Query**
- JSON data returned by Alexa APIs
- SQLite files for Alexa app

**Request parsing**
- Insert saved data info.
Visualization and Evaluation (1/2)
Visualization and Evaluation (2/2)

CIFT Dashboard for
TIMELINE table

Record Count (Total)  Record Count
3,177  1,552
Total records  Unique records

Source-Source Type Distribution Chart

Timeline Table

<table>
<thead>
<tr>
<th>Datetime</th>
<th>Source Type</th>
<th>Short</th>
<th>Desc</th>
<th>Notes</th>
<th>Extra</th>
<th>Count of Duplicated Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016-11-18</td>
<td>Activity</td>
<td>History</td>
<td>History</td>
<td>alexa</td>
<td>SUCCESS</td>
<td>1</td>
</tr>
<tr>
<td>2016-11-18</td>
<td>Activity</td>
<td>History</td>
<td>History</td>
<td>hi</td>
<td>SUCCESS</td>
<td>1</td>
</tr>
<tr>
<td>2016-11-18</td>
<td>Home</td>
<td>TextCard</td>
<td>Hi</td>
<td>Alexa heard: &quot;Hi!&quot;</td>
<td>Alexa's answer: &quot;Hello,&quot;</td>
<td>1</td>
</tr>
</tbody>
</table>
Demo video
Conclusion and Future Works

• Conclusion
  • This paper proposed new approach for Alexa ecosystem
  • We conducted integrated analysis of forensically meaningful data from both systems upon consideration of the target device’s ecosystem

• Future works
  • **Hardware** level of Alexa-enabled devices
  • Performing **memory** forensics for delving into volatile artifacts
  • Digital forensic approaches for **another IoT devices (Google Home)**
  • Implement new component of **CIFT**
  • **Privacy** issue
  • Digital evidence integrity in IoT ecosystem