Digital Forensic Approaches for Amazon Alexa Ecosystem

DFRWS USA 2017

Hyunji Chung, Jungheum Park, Sangjin Lee

Korea University

DFRC Research team

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

Agenda

Intelligent virtual assistants and digital forensics

Digital forensic analysis strategy

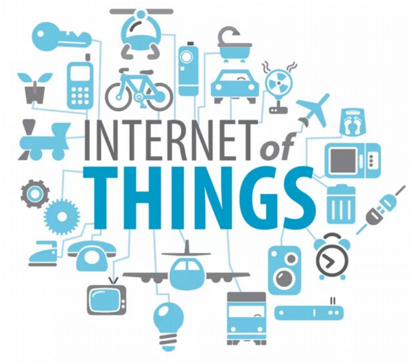
Forensic artifacts on Amazon Alexa ecosystem

Forensic toolkit and visualization

Demo video

Conclusion and future works

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

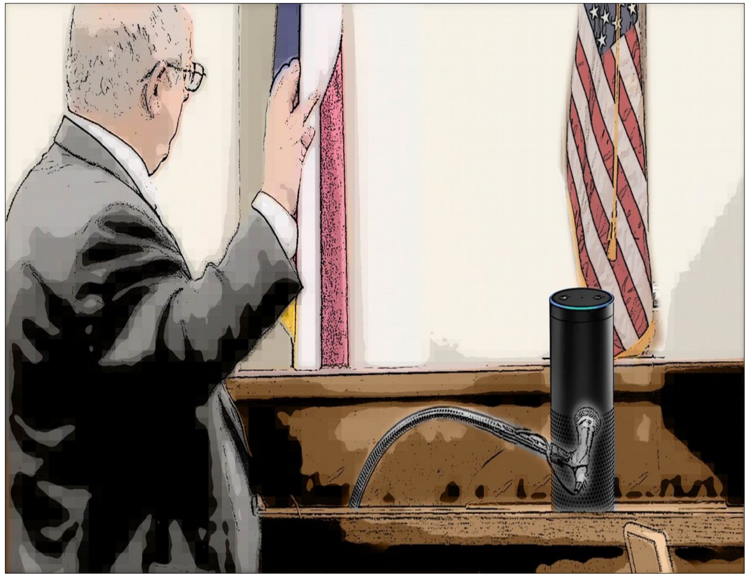
IP camera

Connected car

Intelligent Virtual Assistant and Digital Forensics

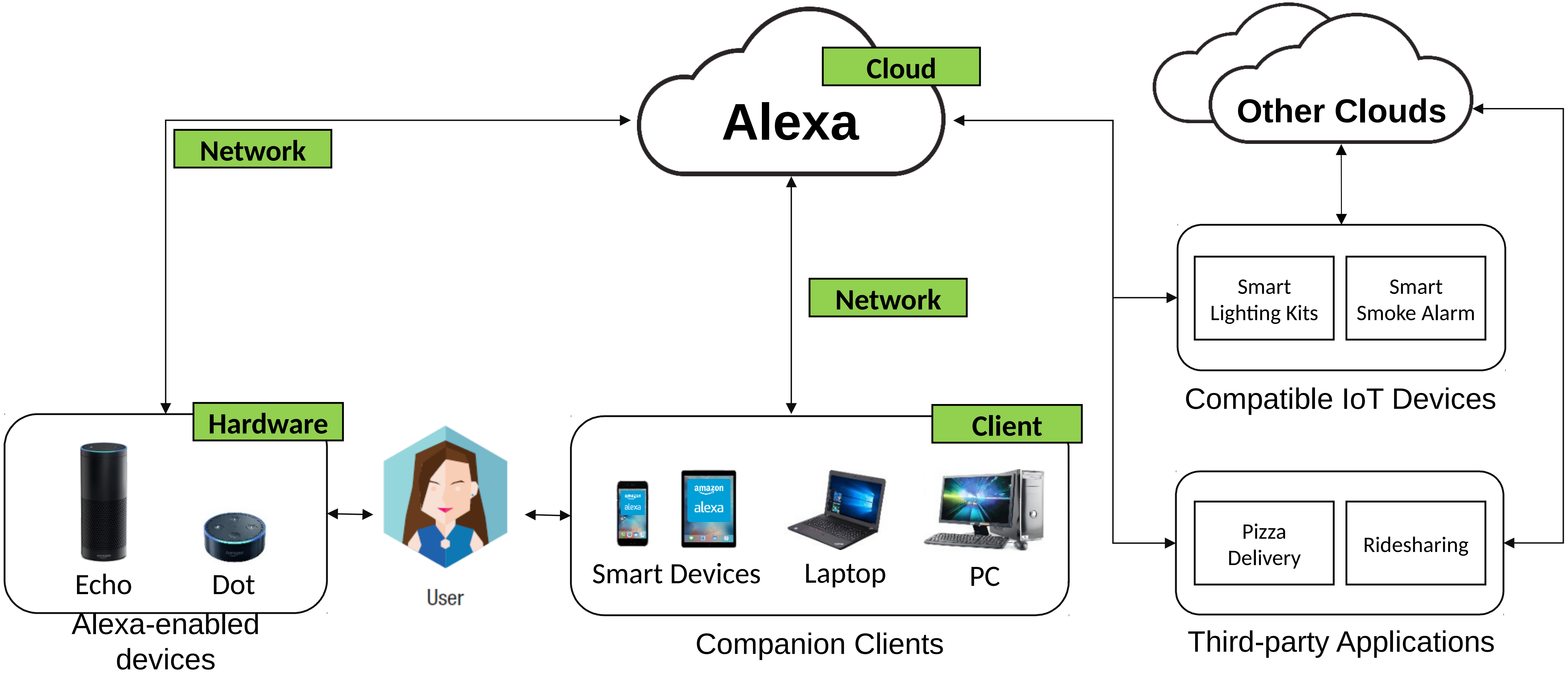
- Intelligent Virtual Assistant and Digital Forensics
  - 25% of households using an intelligent virtual assistant (IVA) will have two or more devices by 2020
  - Amazon Alexa-related environment will become an important source of potential digital evidence (CES 2017)

- 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

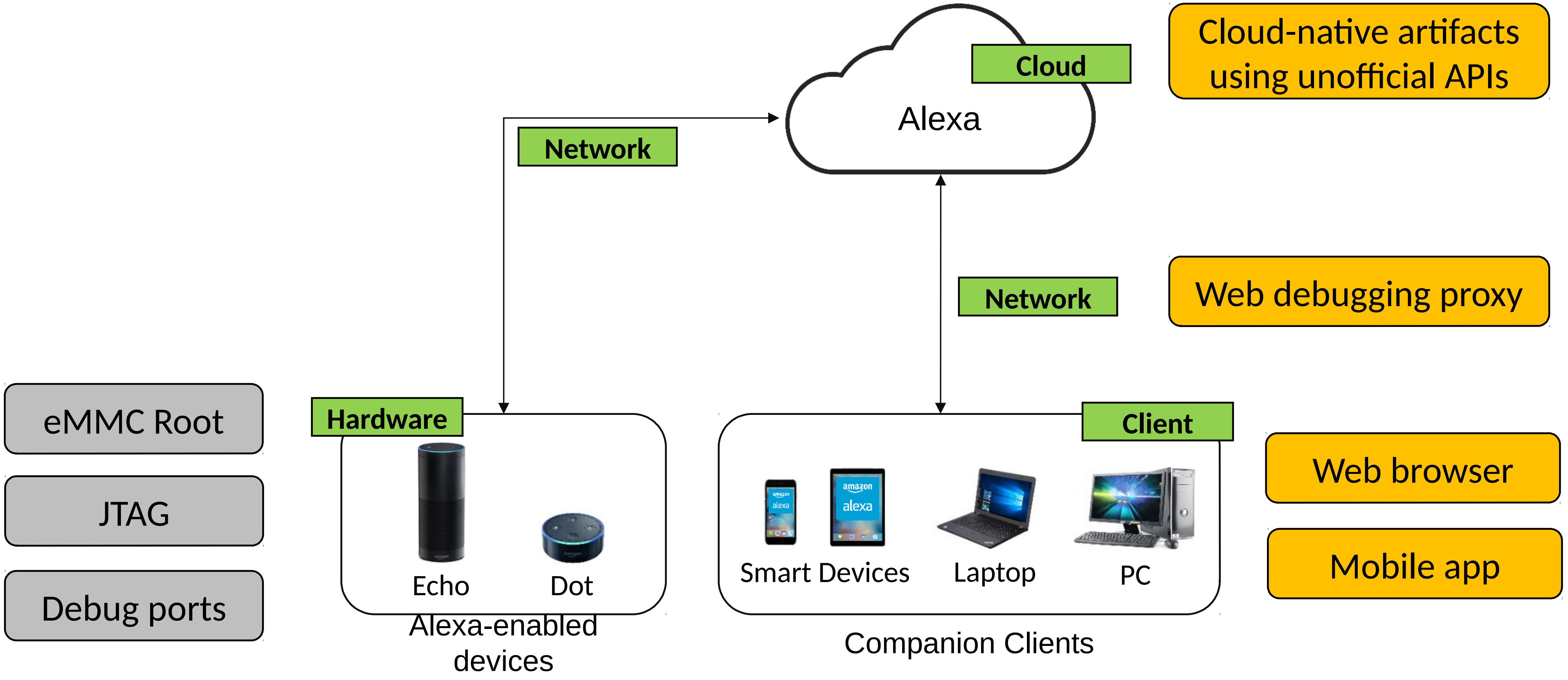


Amazon Alexa and Digital Forensics

- Amazon Alexa Ecosystem



Digital Forensic Analysis Strategy



Alexa

Cloud

Cloud-native artifacts using unofficial APIs

Network

Network

Web debugging proxy

eMMC Root

JTAG

Debug ports

Hardware

Echo

Dot

Alexa-enabled devices

Client

Smart Devices

Laptop

PC

Companion Clients

Web browser

Mobile app

Related works

IoT forensics

Hypothetical IoT crime scenarios
(Oriwoh et al.)

Fundamental challenges for IoT forensics
(Hegarty et al.)

Definition of IoT forensics
(Zawoad et al.)

Forensic framework in the IoT domain
(Kebande et al.)

Cloud forensics

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

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

Forensic Artifacts on Amazon Alexa Ecosystem (2/4)

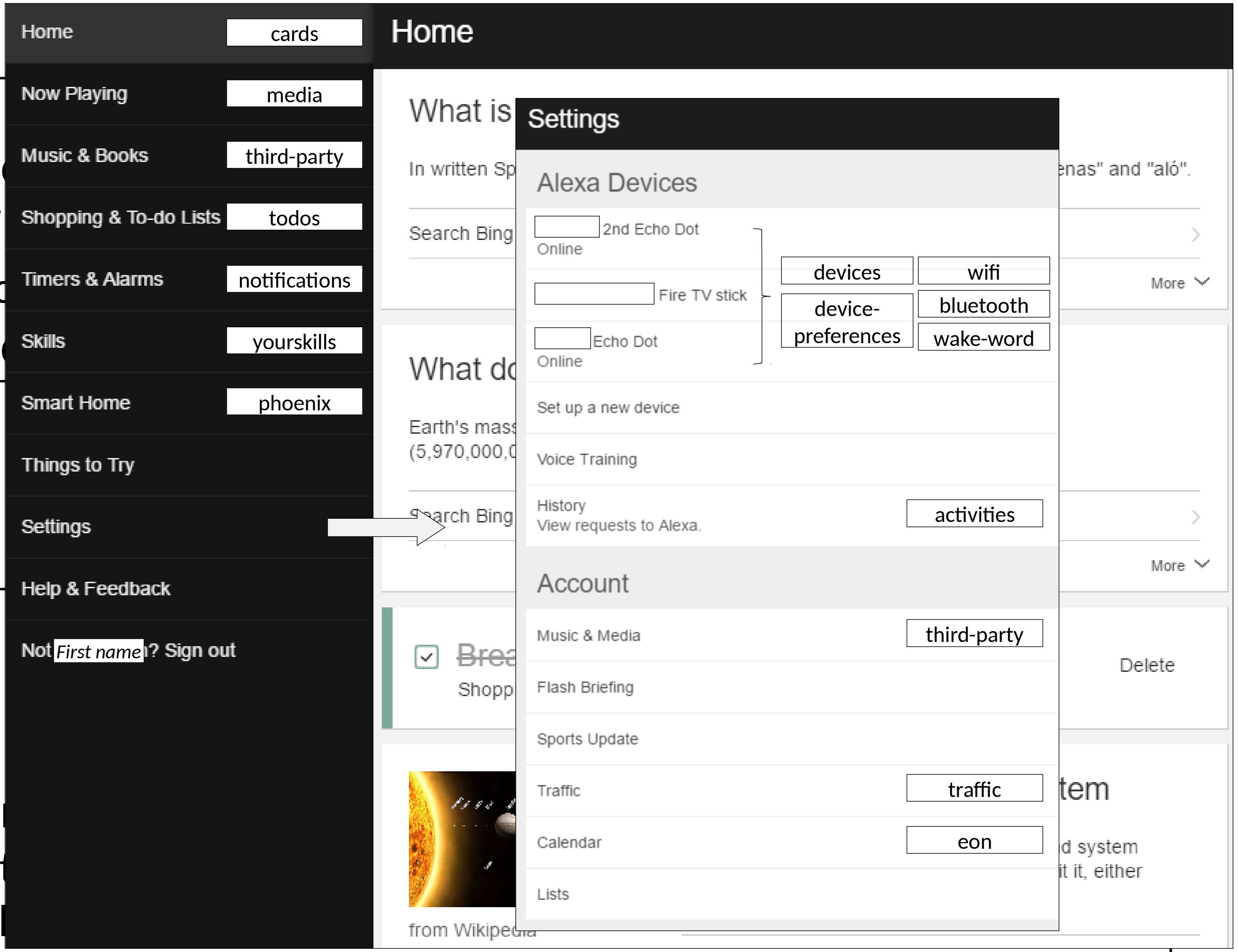
- Cloud native artifacts

Revealing unofficial Alexa APIs

- Understanding the communication proto
- We performed an intensive traffic analysi
  → Most traffic is transferred over encrypted c
- Valid user credential → we can identify a

Alexa's native artifacts

- Seven categories of APIs
  - account, customer setting, Alexa-enable
- Forensically meaningful native artifacts fr
  - registered user accounts, saved Wi-Fi se
  - Alexa-enabled devices, linked Google cal



Forensic Artifacts on Amazon Alexa Ecosystem (3/4)

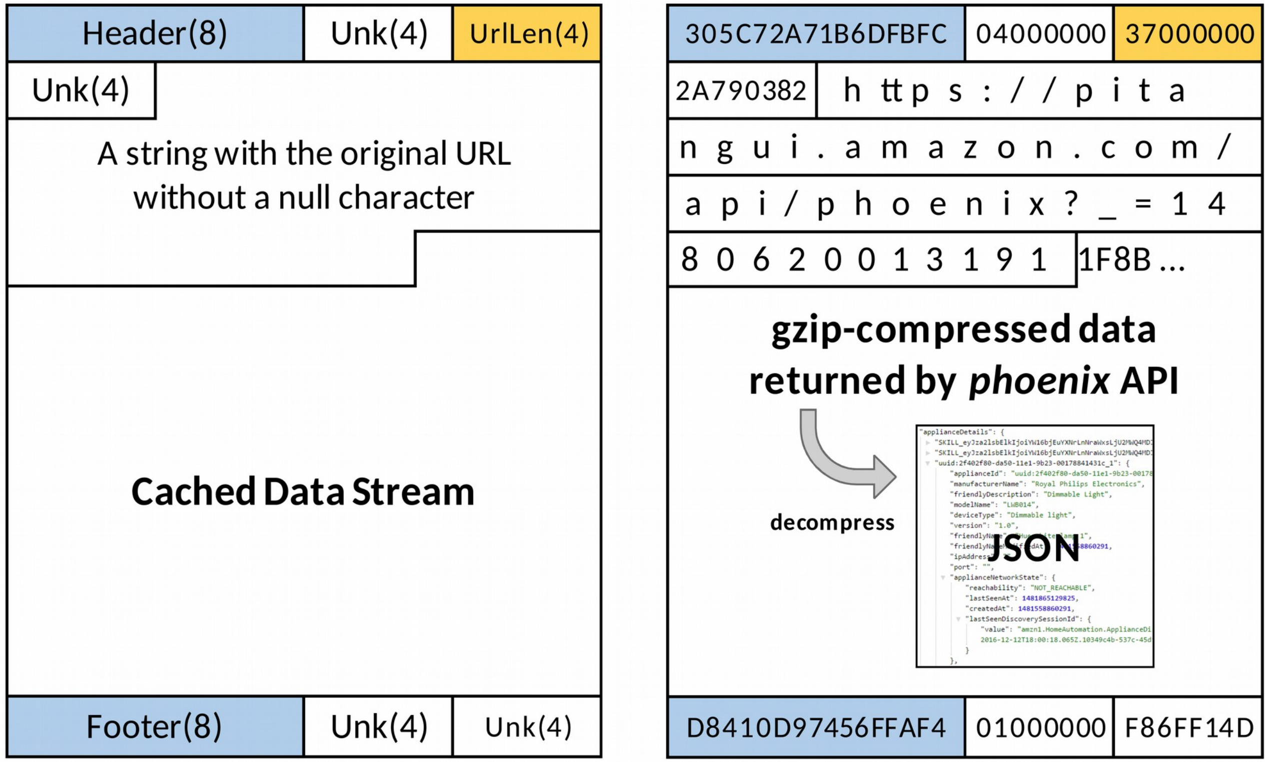
- Client-centric artifacts – Databases of the Alexa mobile app (1/2)

OS	Application	Path	Format	Description
Android 4.4.2	Alexa 1.24.1176.0	/data/data/com.amazon.dee.app/databases/map_data_storage.db	SQLite	**Tokens of an active user**
		/data/data/com.amazon.dee.app/databases/DataStore.db	SQLite	**Todo and shopping list**
		/data/data/com.amazon.dee.app/app_webview/Cache/*	WebView cache	**Cached native artifacts**
iOS 10.1.1	Alexa 1.24.1176.0	[iTunes backup]/com.amazon.echo/Documents/LocalData.sqlite	SQLite	**Todo and shopping list**
OS X 10.10.5	Chrome 55.0.2883.87	~/Library/Caches/Google/Chrome/Default/Cache/	Chrome cache	**Cached native artifacts**
Windows 10	Chrome 55.0.2883.87	%UserProfile%\AppData\Local\Google\Chrome\User Data\Default\Cache\	Chrome cache	**Cached native artifacts**

Forensic Artifacts on Amazon Alexa Ecosystem (4/4)

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

Android WebView cache



Header(8) | Unk(4) | UrlLen(4)

Unk(4)

A string with the original URL without a null character

Cached Data Stream

Footer(8) | Unk(4) | Unk(4)

305C72A71B6DFBFC | 04000000 | 37000000

2A790382 | https://pitanguі.amazon.com/api/phoenix?_=1480620013191

1F8B …

gzip-compressed data returned by *phoenix* API

```
"applianceDetails": {
  ▶ "SKILL_eyJza2lsbElkIjoiYW16bjEuYXNrLnNraWxsLjU2MWQ4MDJ
  ▶ "SKILL_eyJza2lsbElkIjoiYW16bjEuYXNrLnNraWxsLjU2MWQ4MDJ
  ▼ "uuid:2f402f80-da50-11e1-9b23-00178841431c_1": {
      "applianceId": "uuid:2f402f80-da50-11e1-9b23-00178
      "manufacturerName": "Royal Philips Electronics",
      "friendlyDescription": "Dimmable Light",
      "modelName": "LWB014",
      "deviceType": "Dimmable light",
      "version": "1.0",
      "friendlyName": "Hue white lamp 1",
      "friendlyNameModifiedAt": 1481558860291,
      "ipAddress": "",
      "port": "",
    ▼ "applianceNetworkState": {
        "reachability": "NOT_REACHABLE",
        "lastSeenAt": 1481865129825,
        "createdAt": 1481558860291,
      ▼ "lastSeenDiscoverySessionId": {
          "value": "amzn1.HomeAutomation.ApplianceDi
          2016-12-12T18:00:18.065Z.10349c4b-537c-45d
        }
    },
```

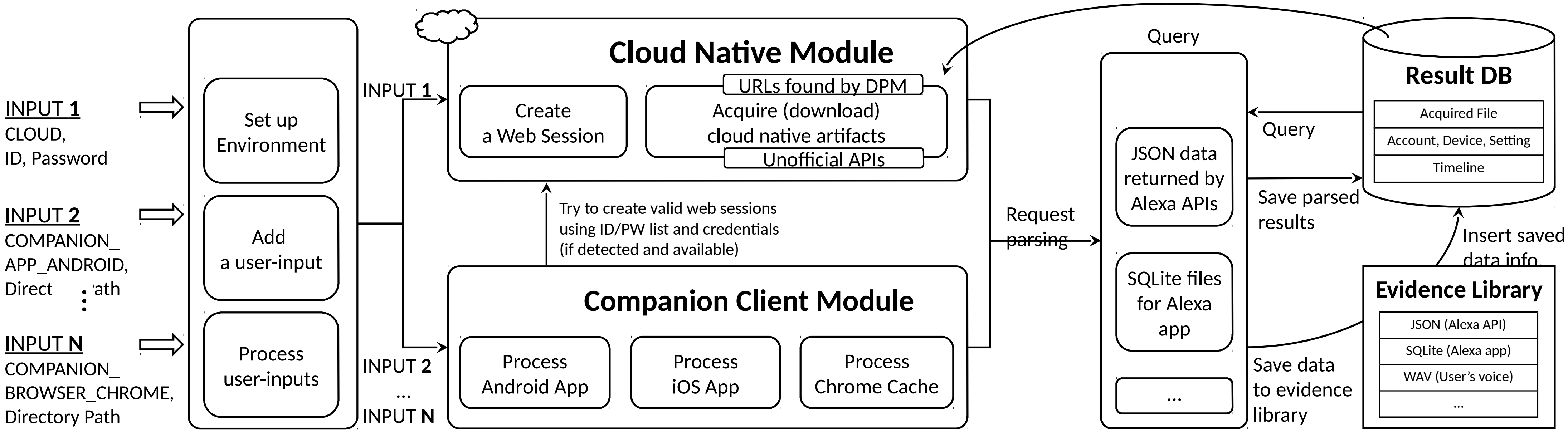
D8410D97456FFAF4 | 01000000 | F86FF14D

(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,
Direct ⋮ ath

INPUT **N**
COMPANION_
BROWSER_CHROME,
Directory Path

Set up Environment

Add a user-input

Process user-inputs

INPUT **1**

INPUT **2**
…
INPUT **N**

Cloud Native Module

Create a Web Session

URLs found by DPM

Acquire (download) cloud native artifacts

Unofficial APIs

Try to create valid web sessions
using ID/PW list and credentials
(if detected and available)

Companion Client Module

Process Android App

Process iOS App

Process Chrome Cache

Request parsing

Query

JSON data returned by Alexa APIs

SQLite files for Alexa app

…

Query

Save parsed results

Save data to evidence library

Result DB

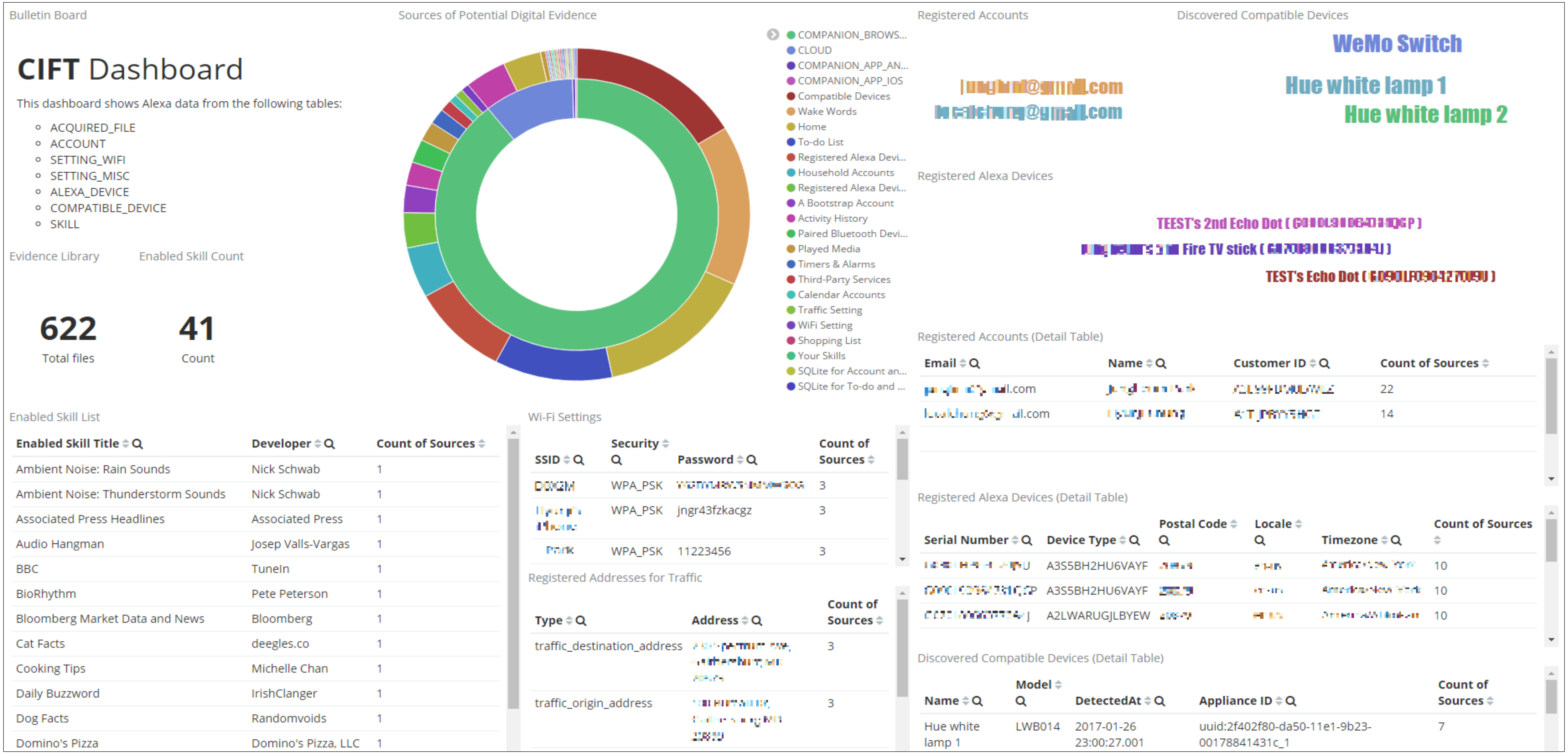
Acquired File
Account, Device, Setting
Timeline

Insert saved data info.

Evidence Library

JSON (Alexa API)
SQLite (Alexa app)
WAV (User's voice)
…

Visualization and Evaluation (1/2)



Bulletin Board

CIFT Dashboard

This dashboard shows Alexa data from the following tables:

- ACQUIRED_FILE
- ACCOUNT
- SETTING_WIFI
- SETTING_MISC
- ALEXA_DEVICE
- COMPATIBLE_DEVICE
- SKILL

Evidence Library

622

Total files

Enabled Skill Count

41

Count

Sources of Potential Digital Evidence

Registered Accounts

[redacted]
[redacted]

Discovered Compatible Devices

Registered Alexa Devices

TEEST's 2nd Echo Dot ([redacted])
[redacted] Fire TV stick ([redacted])
TEST's Echo Dot ([redacted])

Enabled Skill List

Enabled Skill Title	Developer	Count of Sources
Ambient Noise: Rain Sounds	Nick Schwab	1
Ambient Noise: Thunderstorm Sounds	Nick Schwab	1
Associated Press Headlines	Associated Press	1
Audio Hangman	Josep Valls-Vargas	1
BBC	TuneIn	1
BioRhythm	Pete Peterson	1
Bloomberg Market Data and News	Bloomberg	1
Cat Facts	deegles.co	1
Cooking Tips	Michelle Chan	1
Daily Buzzword	IrishClanger	1
Dog Facts	Randomvoids	1
Domino's Pizza	Domino's Pizza, LLC	1

Wi-Fi Settings

SSID	Security	Password	Count of Sources
[redacted]	WPA_PSK	[redacted]	3
[redacted]	WPA_PSK	jngr43fzkacgz	3
[redacted]	WPA_PSK	11223456	3

Registered Addresses for Traffic

Type	Address	Count of Sources
traffic_destination_address	[redacted]	3
traffic_origin_address	[redacted]	3

Registered Accounts (Detail Table)

Email	Name	Customer ID	Count of Sources
[redacted]	[redacted]	[redacted]	22
[redacted]	[redacted]	[redacted]	14

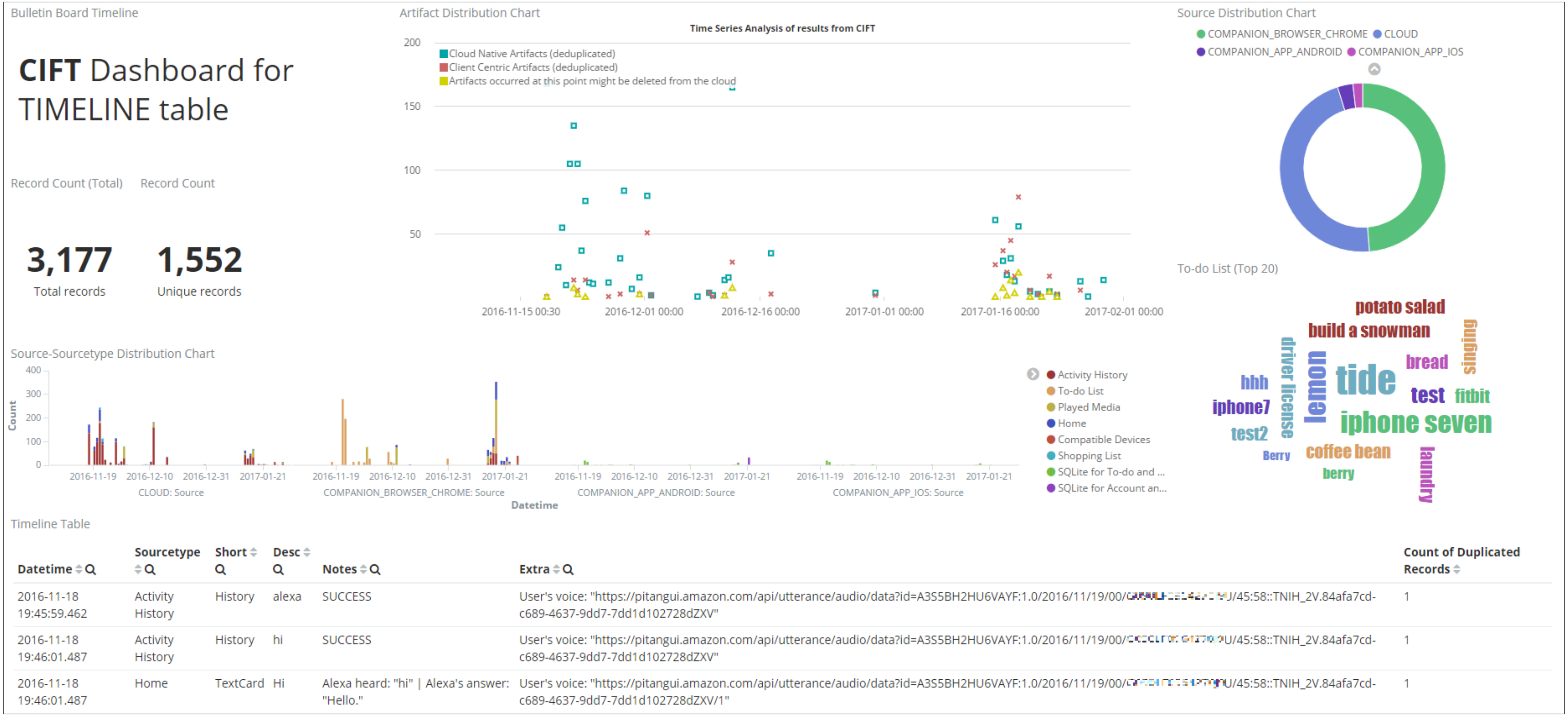
Registered Alexa Devices (Detail Table)

Serial Number	Device Type	Postal Code	Locale	Timezone	Count of Sources
[redacted]	A3S5BH2HU6VAYF	[redacted]	[redacted]	[redacted]	10
[redacted]	A3S5BH2HU6VAYF	[redacted]	[redacted]	[redacted]	10
[redacted]	A2LWARUGJLBYEW	[redacted]	[redacted]	[redacted]	10

Discovered Compatible Devices (Detail Table)

Name	Model	DetectedAt	Appliance ID	Count of Sources
Hue white lamp 1	LWB014	2017-01-26 23:00:27.001	uuid:2f402f80-da50-11e1-9b23-00178841431c_1	7

Visualization and Evaluation (2/2)



Bulletin Board Timeline

CIFT Dashboard for TIMELINE table

Record Count (Total) Record Count

3,177
Total records

1,552
Unique records

Artifact Distribution Chart

Time Series Analysis of results from CIFT
200
150
100
50
🟩 Cloud Native Artifacts (deduplicated)
🟥 Client Centric Artifacts (deduplicated)
🟨 Artifacts occurred at this point might be deleted from the cloud
2016-11-15 00:30
2016-12-01 00:00
2016-12-16 00:00
2017-01-01 00:00
2017-01-16 00:00
2017-02-01 00:00

Source Distribution Chart

● COMPANION_BROWSER_CHROME ● CLOUD
● COMPANION_APP_ANDROID ● COMPANION_APP_IOS

Source-Sourcetype Distribution Chart

400
300
200
100
0
Count
2016-11-19
2016-12-10
2016-12-31
2017-01-21
CLOUD: Source
COMPANION_BROWSER_CHROME: Source
COMPANION_APP_ANDROID: Source
COMPANION_APP_IOS: Source
Datetime
● Activity History
● To-do List
● Played Media
● Home
● Compatible Devices
● Shopping List
● SQLite for To-do and ...
● SQLite for Account an...

To-do List (Top 20)

potato salad
build a snowman
singing
driver license
lemon
hhh
tide
bread
test
fitbit
iphone7
iphone seven
test2
Berry
coffee bean
berry
laundry

Timeline Table

Datetime ⬍ 🔍	Sourcetype ⬍ 🔍	Short ⬍ 🔍	Desc ⬍ 🔍	Notes ⬍ 🔍	Extra ⬍ 🔍	Count of Duplicated Records ⬍
2016-11-18 19:45:59.462	Activity History	History	alexa	SUCCESS	User's voice: "https://pitangui.amazon.com/api/utterance/audio/data?id=A3S5BH2HU6VAYF:1.0/2016/11/19/00/[redacted]U/45:58::TNIH_2V.84afa7cd-c689-4637-9dd7-7dd1d102728dZXV"	1
2016-11-18 19:46:01.487	Activity History	History	hi	SUCCESS	User's voice: "https://pitangui.amazon.com/api/utterance/audio/data?id=A3S5BH2HU6VAYF:1.0/2016/11/19/00/[redacted]U/45:58::TNIH_2V.84afa7cd-c689-4637-9dd7-7dd1d102728dZXV"	1
2016-11-18 19:46:01.487	Home	TextCard	Hi	Alexa heard: "hi" | Alexa's answer: "Hello."	User's voice: "https://pitangui.amazon.com/api/utterance/audio/data?id=A3S5BH2HU6VAYF:1.0/2016/11/19/00/[redacted]U/45:58::TNIH_2V.84afa7cd-c689-4637-9dd7-7dd1d102728dZXV/1"	1

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

Q & A

localchung@gmail.com

https://hyunjichung.github.io