IoT Forensics: Amazon Echo as a Use Case

Shancang Li, Kim-Kwang Raymond Choo *Senior Member, IEEE,* Qindong Sun, William J. Buchanan, and Jiuxin Cao

Abstract—Internet of Things (IoT) are increasingly common in our society, and can be found in applications such as battlefields and national security. These devices can also be targeted by attackers and hence, they are a valuable source in digital forensic investigations. In addition, incriminating evidence may be stored on an IoT device (e.g. Amazon Echo in a home environment and Fitbit worn by the victim or an accused person). In comparison to the IoT security and privacy literature, however, IoT forensics is relatively under-studied. IoT forensics is also challenging in practice, particularly due to the complexity, diversity, and heterogeneity of IoT systems. In this paper, we present an IoT based forensic model that supports the identification, acquisition, analysis, and presentation of potential artifacts of forensic interest from IoT devices and the underpinning infrastructure. Specifically, we use the popular Amazon Echo as a use case to demonstrate how our proposed model can be used to guide forensics analysis of IoT devices.

Index Terms—Digital Forensics, Internet of Things, IoT Forensics, Amazon Echo Forensics

✦

1 INTRODUCTION

IN an Internet of Things (IoT) setting, the number of smart devices connected to the Internet can range from a few to billions. Such devices are often able to sense their environment (e.g. temperature, humidity and wind speed) as well as interconnecting and communicating with each other [1], [2], [3]. According to Juniper research [4], more than 20.4 billion smart devices will be connected to IoT by the year 2020, generating approximately £134 billion annually by 2022 for the IoT cyber security industry. This is telling of the IoT trend in our society, which is also evident by IoT being extended to sectors such as battlefields and military (e.g. Internet of Battlefield Things and Internet of Military Things)[1].

In 2017, it was reported that users of Bose headphones were being spied upon without their consent [5]. Specifically, a plaintiff filed a complaint against Bose for their Bose Connect application, which allegedly collects data on the music and audio books, that their users listen to, and sends the collected information to a third-party data miner (Segment.io). In the same year, Vizio [6], a Smart TV manufacturer, was also allegedly monitoring over 11 million smart TVs, where user data were being sent to other third parties without user consent [6]. Specifically, they monitored the pixels displayed on the TV screen and matched these to

movies stored on a database. This technique is known as automatic content recognition (ACR). Vizio was subsequently fined a total of $USD$$2.2$ Million by the US Federal Trade Commission, and was also ordered not to track their users [6]. In addition, the organization was ordered to delete all their existing data relating to this incident (e.g. near-by access point details, postal codes, and the Internet protocol address (IP Address) of the local network, and implement a privacy policy.

In general, an IoT system consists of IoT devices, IoT infrastructures, services and applications, and interface to other applications or services, which can be organized into four layers as shown in Figure 1:

1) Sensing layer, which includes sensing devices to sense and acquire information, such as smart sensors, radiofrequency identification (RFID), and end components of IoT;
2) Network layer, which is the infrastructure to support connectivity to Internet and other devices;
3) Service layer, which provides and manages services to users or other applications; and
4) Application-interface layer, which provides interface to users or other services.

It can be expected that the increasing popularity and pervasivenesss of IoT devices will be targeted by attackers to compromise our systems or exfiltrate our data, and gain a competitive advantage. In other words, any IoT device such as a 3D printer, a smart switch or a smart bulb in a smart home environment can potentially be compromised to gain access to the smart devices or the user's personal data [1], [3], [7], [8]. In 2016, for example, distributed denial of service (DDoS) attacks targeting the Domain Name System (DNS) provider, Dyn, was carried out by a botnet comprising a large number of 2.5 million compromised IoT devices (e.g. IP camera, smart printers, and home WiFi gateway) [2], [9]. In the past two years, a large number of vulnerability scans targeting IoT devices have been reported. It was also reported that nearly 65% of today's IoT devices are insecure [10]. This is not surprising due to the challenges in designing efficient security and privacy solutions. In other words, security and privacy solutions designed for IoT

- *S. Li is with the Department of Computer Science and Creative Technologies, University of the West of England, Bristol BS16 1QY, UK.*
  E-mail: shancang.li@uwe.ac.uk
- *K.-K. R. Choo is with the Department of Information Systems and Cyber Security, University of Texas at San Antonio, San Antonio, TX 78249, USA*
  E-mail: raymond.choo@fulbrightmail.org
- *Qindong Sun is with the Department of Computer Science at Xi'an University of Technology, Xi'an 710048, China*
  E-mail: sqd@xaut.edu.cn
- *William J. Buchanan is with the School of Computing, Edinburgh Napier University, Edinburgh EH10 5DT, UK*
  E-mail: w.buchanan@napier.ac.uk.
- *Jiuxin Cao is with the School of Computer Science and Engineering, Southeast University, Nanjing, China*
  E-mail: jx.cao@seu.edu.cn.

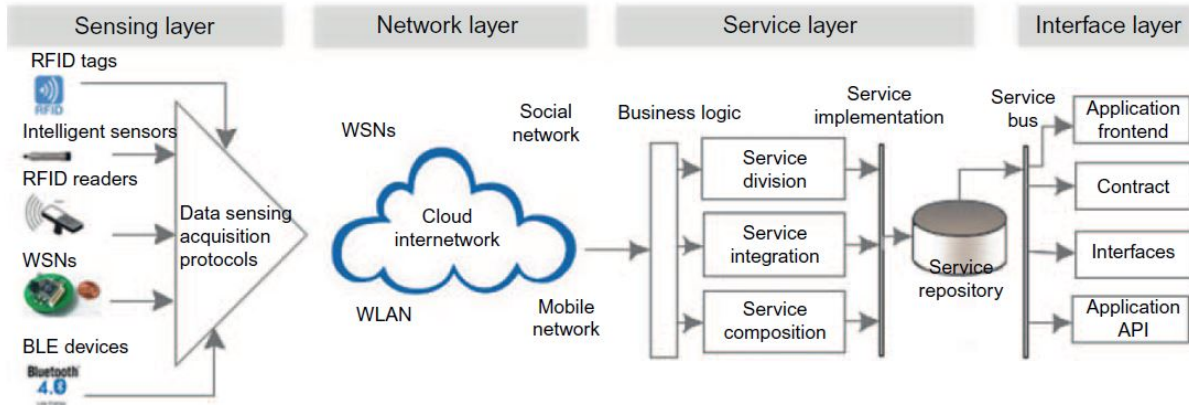1. https://www.arl.army.mil/www/default.cfm?page=3050 (last accessed February 12th, 2019)

Fig. 1. General IoT Architecture

devices will have to take into consideration the interoperability and complex ecosystems, as well as the computational limitations in IoT devices.

Hence, IoT devices are likely to be sources of evidence in a cyber security investigation (e.g. investigation of a DDoS attack) [11]. Unlike conventional digital forensics (e.g. mobile device forensics), the diversity in IoT devices (e.g. 3D printers, roadside units in a smart transportation system, smart healthcare devices in a hospital, and smart military uniforms) and the different evidence and privacy regulations compound the challenges of such investigations [12]. Some of these challenges are as follows:

- *Identification*: identification of potential evidence in IoT environment can be challenging, particularly if the investigators are not familiar with the types of IoT devices present as well as the underpinning infrastructure;
- *Preservation*: once the potential source of evidence is identified, then the question is how can we acquire and preserve the evidence from the IoT devices, companion application, IoT services, networks in the IoT infrastructure, and so on, in a forensically sound manner; and
- *Analysis*: depending on the format that the evidence is acquired, analysis of the acquired evidence may be challenging. We also have to ensure that the analysis takes into consideration data provenance and the interaction between IoT and cloud servers that facilitates the aggregation and processing of data from the IoT.

The following major contributions are presented in this paper:

1) We propose an IoT based forensic analysis model, which supports the identification, acquisition, analysis, and presentation of potential artifacts of forensic interest from IoT devices and the underpinning infrastructure;
2) We address IoT devices forensic investigation processes from the forensic perspective, in which each IoT devices are expected to provide important forensic artifacts;
3) We analyse forensic artifacts retrieved from the popular Amazon Echo as a use case to demonstrate how our proposed model can be used to guide forensics analysis of IoT devices.

In this paper we present an IoT forensic model (see Section 3) and demonstrate how it can be used to guide the investigation of IoT devices, using Amazon Echo as a case study in Section 4. In the next section, we will briefly discuss related literature.

2 RELATED LITERATURE

In recent years, IoT forensic has attracted attention from the forensic community [13], [14], [15], [16], for example in wearable devices [17], smart vehicles [18], smart home devices [19], and so on. Approaches may vary between the nature and type of digital forensic investigation. For example, at the IoT network layer, network forensics tools or methods are generally applied. We refer interested readers to the work of *Caviglion et al.*, who reviewed popular digital techniques in network forensics, reverse engineering, and so on, as well as the prevalent data storage formats and files systems [20]. Key challenges were also briefly discussed [20]. In another related work [14], the authors categorized IoT forensics into three zones: IoT zone, network zone, and cloud zone, where each zone consists of different areas and forensics analysis activities.

In [21], the authors presented an automated forensic management system (FEMS) that was designed to collect data from a three-layered architecture, namely: perception, network, and application layers. However, in dynamic IoT networks, it is difficult for FEMS to investigate all states of the IoT devices. *Zawoad et al.* proposed a forensic-aware IoT (FAIoT) model which allows collected evidence to be stored in a secure evidence repository server [10]. In [15], Orlando *et al.* described the methods to investigate the device's hardware and the relevant system (e.g. operating system, boot loader, remote installation and communication system). In addition, a detailed security measurement for IoT devices was provided. In [22], a general IoT forensics framework was proposed, comprising a forensic state acquisition (FSAIoT), and a centralized forensic state Acquisition (FSAC) to classify the evidence acquisition of IoT devices into three modes (i.e. controller to IoT devices, controller to cloud, and controller to controller) [23].

There have also been research efforts in smart home devices and the forensic of such devices. For example, Amazon Echo is increasingly used as the voice controller hub of smart sensors and devices, which plays a centric role in bridging different smart home devices and the amazon cloud server. The Amazon Echo is activated by wake words like 'Alexa', but must also constantly listen for the wake-up command, and clearly this is a potential evidence source [16], [19], [24]. For example, Chung *et al.* [25] explained how companion clients (i.e. devices used to send and capture commands and responses from intelligent home assistants, such as Alexa) can also be a source of evidence.
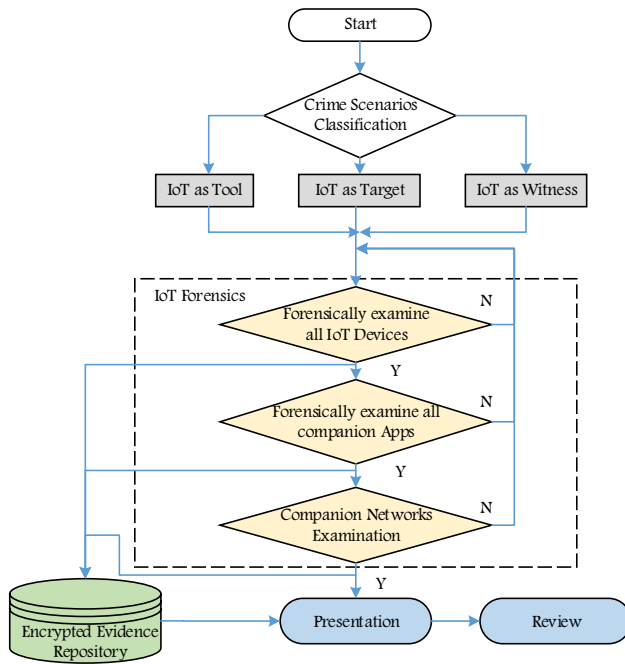
Fig. 2. Proposed IoT Forensic Model

A number of device fingerprinting techniques have also been developed, which can be used for the investigation of IoT devices. For example, sensor pattern noise (SPN) can be used to identify the source device that has acquired a digital image or video, and this is relevant for the investigation of IoT devices that have a image or video acquisition capability (e.g. unmanned aerial vehicles). In SPN based image forensic analysis, as the most dominant part of SPN the photo response non-uniformity (PRNU) noise can be extracted from an image to build image fingerprint and camera fingerprint, which has been widely used in image origin identification and image forgery detection. Flicker forensics can also allow an investigator to identify an IoT device by analyzing the flicker signal and associate the parameters with some internal characteristics of the particular device [26].

In the next section, we will address our IoT forensic model.

3 PROPOSED IoT FORENSIC MODEL

When we conduct an IoT forensic investigation, we have to consider the sources of evidence other than the actual IoT devices, for example, the sensing, network, service and interface layers (see Figure 1).

Similar to conventional digital forensics, IoT forensics mainly consists of the following four stages: *identification, preservation, analysis, and presentation* [27].

1) In the identification stage, the focus should be on IoT devices (e.g. sensors and intelligent home assistants such as Amazon Echo), and any related infrastructure (e.g. routers).

2) In the preservation stage, we may require specialized / customized tools to acquire data from (proprietary) hardware and applications.

3) In the analysis stage, customized forensic tools may be required to analyze data from certain devices, other than the typical commercial forensic tools (e.g. EnCase and FTK). Both EnCase and FTK are commonly used forensics tools that can be used in digital security, security investigation, and e-discovery.

4) In the presentation stage, forensic investigators will need to detail the findings and be able to articulate the analysis, findings and their implications in a court of law. Meanwhile, the evidence items should be presented with their original format.

Building on the typical four-stage digital forensic process, we present an IoT forensic model – see Figure 2. Specifically, our model starts from an *offense classification* stage, where the roles of IoT are classified into *IoT as a target, IoT as a tool*, and *IoT as a witness*. Then, each related device and the companion apps are examined using the above four-stage process. In addition, all acquired forensic artifacts are stored in an encrypted evidence repository.

3.1 Offense Classification

Due to the diversity of devices and heterogeneity of networks in an IoT setting, it can be challenging to identify all sources of evidence and collect all relevant forensic artifacts in a timely fashion, especially if third parties or remote servers (e.g., websites and cloud servers) are involved. Firstly, to effectively identify the devices for an investigation, it is important to consider the nature of the offense (e.g. a serious and organized crime type will generally mean that more resources should be spent on the case), data acquisition methods, and relevant laws (e.g. what are the elements of proof) and regulations. In general, the IoT-related crimes can be group into three classes [28]:

(1) IoT device as a target (e.g. cyberattacks where vulnerabilities in IoT devices are exploited). Many IoT devices are resource limited (e.g., computation capabilities, storage space, power supply, etc.), and most security tools/packages cannot be installed in IoT devices, which make them an easy target for cyber attacks.

(2) IoT device as a tool, IoT devices can be used by forensic investigators as tools to identify, collect, analyse, or even present evidences in digital investigation. For example, a compromised IoT device is been used to facilitate other malicious activities such as a botnet attack.

(3) IoT device as a witness (e.g. data stored in the IoT device can directly implicate an individual accused of a crime), in which IoT devices are able to identify, collect, and preserve evidential data for forensic investigation. One prominent example involved the Amazon Echo, where an Arkansas man was accused of killing his friend. The prosecutor then sought recordings from the defendant's Amazon Echo to be used as evidence [29]. IoT as a witness will likely happen again in the future frequently because IoT devices are now an integral part of our daily life.

Figure 3 shows the workflow of IoT device identification, in which an IoT device will be examined using the appropriate approach.

3.2 IoT Device Identification

In this stage, we seek to answer the following questions:

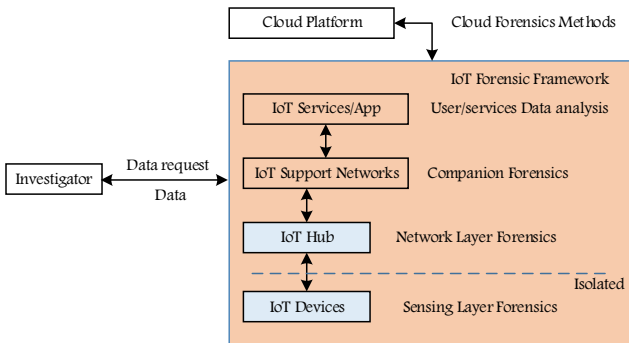- What was/were available at the event/crime scene or a remote site?

Fig. 3. IoT device identification procedure

Fig. 4. IoT Devices identification in IoT Forensics

- Who and what was/were there when the event/crime occurred?
- What are the constraints in collecting the required evidence?
- What is the minimum set of evidence required to support the elements of proof for this specific offense?

A six-step IoT device identification method is presented in Figure 4.

- Define device space, to identify the devices relating to the specific case;
- Establish the device lifecycle, to identify the time span for the device examination;
- Establish access, to identify the accessibility of the devices, including *confidentiality, authentication, authorization*, and so on;
- Define data categories, to define the data category that the device can provide;
- Network access control, to identify the connectivity of the networks relating to the device and isolate the device from the connections;
- Identify the access to devices, this stage summarizes previous steps and establish the availability of the device for investigators.

Despite the diversity of IoT device manufacturers, IoT devices share some similar features and capability. In general, an IoT device consists of a processer or micro-controller, read-only memory (ROM), random access memory (RAM), communication module (Bluetooth, Wireless, Zigbee, *etc.*), and data input/output interfaces. To record the collected or generated data, an IoT device may be equipped with built-in secure digital (SD) memory to support removable memory. Software features of an IoT device include operating systems (some simple IoT devices may only run very simple code without an operating system), middleware, file system, and applications. Many IoT devices do not have a specific file system and in this case, the investigator may need undertake further research, for example how to leverage the application software development kit (SDK) to obtain more information.

Conventional digital forensic tools, such as DD, EnCase, FTK Imager, and SIFT, may also be useful in some cases. In IoT forensics, the data extraction tools/methods can be classified into five levels, namely: manual, logical, hex dumping/JTAG, chip-off, and micro-read [30]. For IoT devices that are not supported by
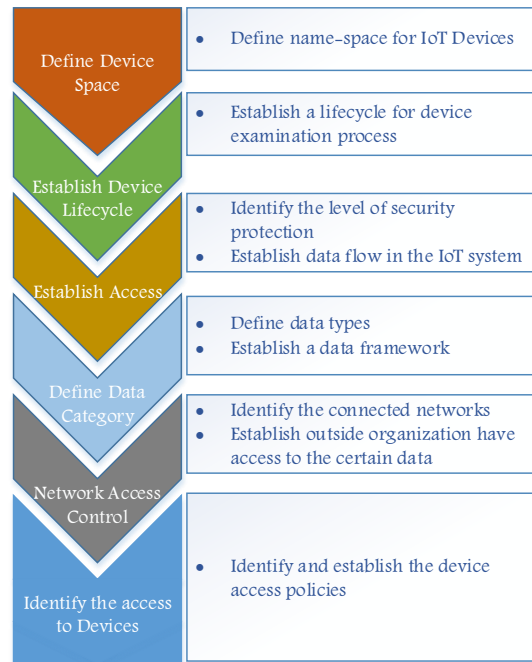
TABLE 1
Example of Hardware Characterization for IoT Devices

Hardware Features	Capabilities	Evidence Sources
Processor	48MHz to 2 GHz	Cache
Microcontroller	Less than 52MHz	RAM and FLASH
Memory	5MB to 128GB	Chip Memory and
Card slots	MicroSD, MicroSD	SD-based Card
Display	LCD to HD	–
Camera	Still, Video	DF/SD Card, ex-Memory
Interfaces	SPI, I2C	built-in RAM/SRAM
JTAG	JTAG Scanner	–
Input Interface	RS232 to Keyboards	–
Voice Input	Voice recognition	EEPROM, ex-storage
Positioning	GPS receiver	External storage
Wireless	IrDA, BT, WiFi, NFC	–
Battery	Li-Ion Polymer	–

existing forensic tools, the investigator could also consider seeking the cooperation of the device owner, reviewing seized material, seeking the assistance of the service provider (e.g. Amazon in the case of Amazon Echo), and so on.

3.3 Evidence Preservation

Tables 1 and 2 show the potential avenues for data preservation, and in this paper we will focus on memory forensics. Specifically, we will focus on (1) extracting data from the memory of a target IoT device; and (2) analyzing the physical memory data (from RAM), page file (or SWAP space) data, etc. Swap space denotes areas on disk used for interchanging contents between main ram and secondary memory, in linux swap is an actual disk partition and in windows machine, the swap space is a pagefile. In digital forensics, Swap file is a rich source of key evidence items, including passwords, sensitive data, encryption keys, etc.
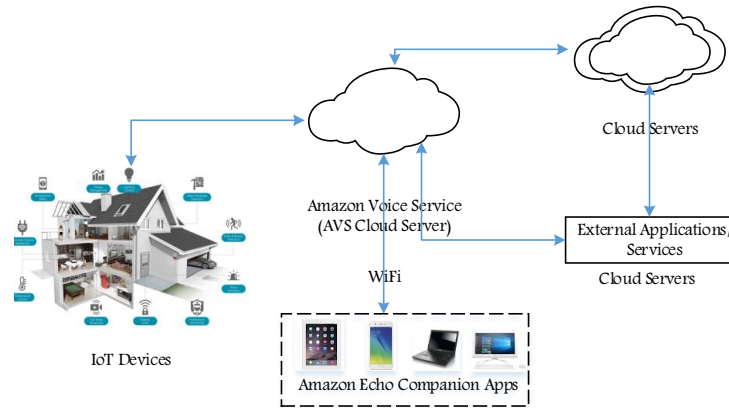
Fig. 5. General IoT Forensic Analysis

TABLE 2
Example of Software Characterization for IoT Devices

Software Features	Capabilities	Evidence Sources
OS	Closed	Android, iOS,etc.
Secure coding	Programming language	RAM and FS
Logs	Events logs, data logs	FS, ex-storage
PIM	Calendar, List	FS, storage
Applications	application data	Cache, RAM, FS, ex-S
Data type	Application-based	Cache, RAM, FS, ex-S
Call	Logs	File systems
Data Processing	depends	Cache, RAM, FS, ex-S
Email	Applications	RAM and FS
Protocols	Communication	Cache, RAM, FS, ex-S
Web	Web browser	Cache, RAM, FS, ex-S
Web	Network services	Cache, RAM, FS, ex-S

Live memory evidence extraction is another major issue in IoT forensic preservation. In resource-constrained IoT devices (e.g. limited computation, storage, energy supply, etc.), volatile memory extraction can often be conducted to extract key evidence stored in the RAM or an ongoing communication session [31]. A number of memory acquisition tools have been developed in the literature, such as the Android based memory subsystem (ashmem) [32], *Android low memory killer* [33], and *memory grab* [34].

However, there are still challenges in live memory acquisition. For example, the *memory protect unit* (MPU) technology only allows specific instructions or code to access the memory. This prevents the forensic investigator from accessing the memory. In addition, anti-forensics (AF) techniques including activities to overwrite data and metadata compound the challenges of memory acquisition. For example, TimeStomp[2] can be used to overwrite NTFS `create, modify, access, and change` timestamps [35].

Also, while a number of tools have been developed for live memory acquisition from computers and laptops (e.g. *Winen, dd, dumpit.exe, winhex, nigilant32, memoryze, readline*), there are limited tools designed for IoT devices.

3.4 IoT Forensic Analysis and Presentation

IoT forensic analysis can be scenario- and device-specific, since IoT systems can have different configurations and settings. For example, in a smart home system as shown in Figure 8, the devices involved may differ from an industry IoT (IIoT) systems. The general approach can include attempts to reconstruct the IoT crime/event scenes. The findings of the analysis also need to be documented and presented, for example to the jury, prosecutors and judges.

4 AMAZON ECHO (PI) FORENSICS

Amazon Echo is a popular intelligent home assistant or 'smart home' IoT hub, which takes voice commands from the users to control itself and other connected IoT devices/sensors (*e.g.* smart lights, smart kettles, smart locks, smart thermostats, and smart doors) [36]. Using the voice recognition technology (i.e. Alexa in the case of Amazon Echo), users can interact with the connected IoT devices using their voice. Clearly, the devices require some sort of Internet connection (*e.g.* WiFi) [16].

In a prior work involving the analysis of Amazon Echo [25], it was reported that the user's history data and interactions with Alexa are stored in the SQLite database and web cache files. The authors analyzed two Amazon Echo Dots, with Android 4.4.2 + Alexa app, iOS 10.1.1 + Alexa app, OS X 10.10.5 + Chrome and Windows 10 + Chrome. For the network analysis, it was determined that most of the communications were encrypted and the JSON format was used for passing parameters. The authors' analysis of the communications revealed undocumented API calls to RESTful Web services. In other words, there are seven categories of data on the device, namely: account, customer setting, Alexa-associated devices, skills and behaviours of user, user activity, and *etc*. The researchers found that most of the data contain UNIX timestamps, which could be used to create timeline of activities within an investigation [25]. Within this applications, the utterance API could be used to download voice files [37].

The location of the client artefacts depends on the access method being used, such as for SQLite databases on iOS and Android, and within Chrome caches for OS X and Windows 10. A summary of these locations is presented in Table 3.

TABLE 3
Location of client artefacts [25]

OS	Application	Path	Format	Description
Android 4.4.2	Alexa 1.24.11760	/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 artificats
iOS 10.1.1	Alexa 1.24.11760	[iTunes backup]com.amazon.echo/Documents/LocalData.sqlite	SQLite	Todo and shopping list
OS X 10.10.5	Chrome 55.0.2883.87	/Library/Cache/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

Name	Date modified	Type	Size
AmazonEchoImage.E01	30/05/2017 09:23	EnCase Evidence F...	1,535,870 KB
AmazonEchoImage.E01.E01	30/05/2017 09:53	Text Document	2 KB
AmazonEchoImage.E01.E02	30/05/2017 09:30	E02 File	1,535,906 KB
AmazonEchoImage.E01.E03	30/05/2017 09:48	E03 File	1,136,293 KB

Fig. 6. Alexa Pi firmware images created using Encase 7.0

- Devices name
- Devices Group
- Serial number
- Device software version
- Device registered to
- Device Location
- Device timezone
- Wake word
- Language
- Temperature Units
- Distance Units

- WiFi Connections: Gateway IP, IP, MAC address, Server address
- Bluetooth Connection: Paired devices

- Calenders
- Lists
- To-do list
- Alarms
- History
- Music
- Books

- Username
- Password
- Host name
- Client version
- Bridge version
- ProductID
- ClientID
- ClientSecret
- Device Reg name
- Bearer token

Fig. 7. Data type related to Amazon Echo

On Android device, the SQLite files are contained in *map_data_storage.db* (token information for the current user, and is deleted when the user signs out) and *DataStore.db*. For iOS device, there is a single file named LocalData.sqlite. While the Android analysis was fairly easy, the iTunes backup protocol had to be used in iOS analysis. The chrome access data was found stored in the data-block-files, which could be possible to rebuild Alexa-related caches into the first HTTP headers, and cached data. This could be useful for determining user behaviors as the stored things (*e.g.* user clicks) can lead to calls to Alexa APIs [37].

In IoT forensics, analyzing embedded files and data with firmware images is an effective way. By connecting the `UART` port in Echo, the boot debug messages can be output to a terminal. In our research, we determine that Echo uses `u-boot` as its boot loader, which is a popular open source bootloader and a number of commands/tools can be used to extract information in the firmware. In this paper, we use the `Alexa Pi` to build an `Echo` over Raspberry Pi Version B, which uses similar firmware with Amazon Echo. In our experiment, we analyze the Alexa Pi over Ubuntu (16.04), the companion app installed on an iPad 4 (iOS 12), and the Alexa Voice Server (AVS). In [38], an AlexaPi is developed to monitor and control devices in smart home environment.

We first use `u-boot` to output the firmware in Alexa Pi, which results in three Encase images (see Figure 11).

4.1 Data Type

We then analyze the data type created, transmitted, processed, and stored on the IoT devices. For an Amazon Echo and the AVS service, we determine that the following (see also Figure 7):

- Device related data include device name, device group, serial number, hardware data, timezone, region, *etc.*
- Connectivity includes connection address, WiFi: Gateway IP, IP, media access control address (MAC address), Server address, Bluetooth address, *etc.*
- User data include data related to the IoT device, such as username/password, language, calenders, and email.
- Application data include Host name, Client version, ProductID, ClientID, ClientSecret, Device Reg name, Bearer token, registered user, *etc.*
- Other data include communication data, specific protocol type, *etc.*

4.2 Alexa Pi Data Acquisition

1) Data acquired from the companion app include device related information, account, and network, as shown in Figure 8. For each IoT device, more detailed data can be extracted using both logical and physical methods, including device name, wifi, device register, serial number, and MAC address. In addition, information such as language and location can also be extracted directly after further analyzing the app.
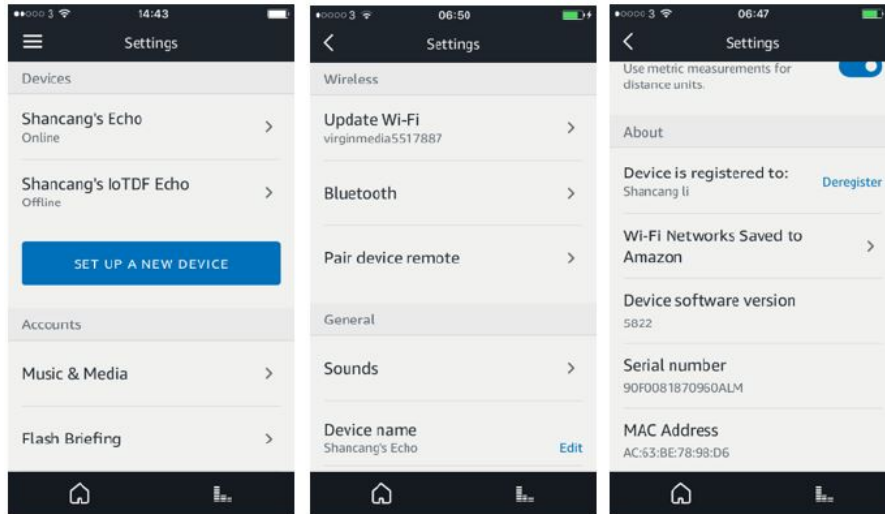
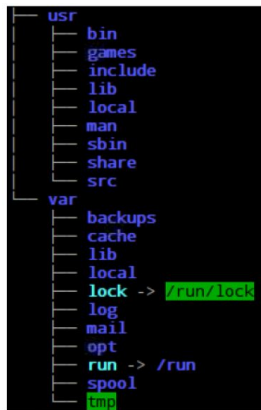Fig. 8. Data extracted from the companion app (partially)

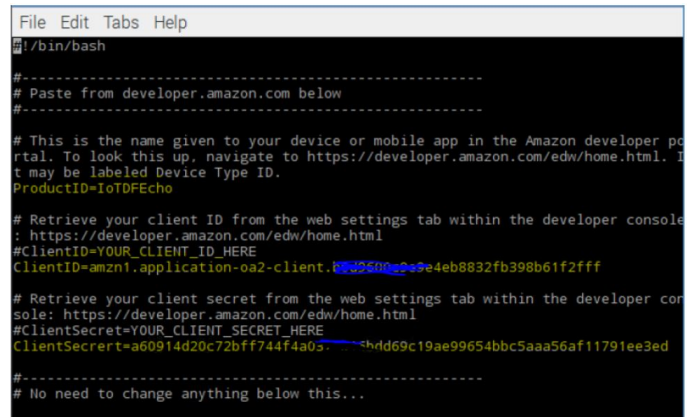Fig. 9. Alexa Pi fireware file system

Fig. 10. Alexa Pi configuration for AVS

The bootloader's command line interface allows raw access to part of the memory areas and Flash integrated circuit (IC). When processing the `bootloader` message via the universal asynchronous receiver/transmitter (UART) port, an investigator can obtain the location of the kernel image and scout the firmware by using the u-boot command-line interface (CLI). In further examination of the file system in the firmware, Debian system information can also extracted as shown in Figure 10.

We use *Zenmap* to locate the IP address as: `192.168.0.10` and MAC address `AC:63:BE:78:98:D6` of an Echo via a ping scan. In more complex investigations, we can also use port pings to find all ports open on the devices. By checking the IP address and MAC address, the investigator can identify other IoT devices that need to be examined.

Through the UART, we can dump the firmware to an image file. In this investigation, it is very difficult to solder the UART to the USB ports. Fortunately, in [39], an Echo image is provided that can be loaded via a raspberry pi, which works fine as an Echo device. We investigate the images on the Raspberry Pi, which contains the information that an Echo has.

Some information to identify the device can be found by investigating the firmware. Amazon requires each Amazon Echo device to provide the productID (also known as Device Type ID), ClientID, and ClientSecret in order to use AVS. The Echo firmware contains several files within the root folder, for example in the automated_install.sh file.

4.3 Examination and Analysis

The two devices' information are shown in Figure 12. For each device, information such as device name, wireless connections, device register, serial number, and MAC address are located and analyzed. For example, the `Setting` section stored within the `Alexa` app contains information that can be used to identify the IoT device. In other words, the investigator can extract the device name, the Wi-Fi that the IoT device had previously connected to, Bluetooth connection information, and paired device. The 'device is registered to' information may also be used to identify the owner of the Echo, for example in collaboration with Amazon. Meanwhile both serial number and MAC address can be used to identify the Echo and other connected devices.

Echo uses an address set in the Alexa companion app, where the location information is used to provide weather forecasting and location-based services. Analysis of location information acquired from companion devices (e.g. Google Maps, Find your device, and weather) can also be corroborated with other analysis.
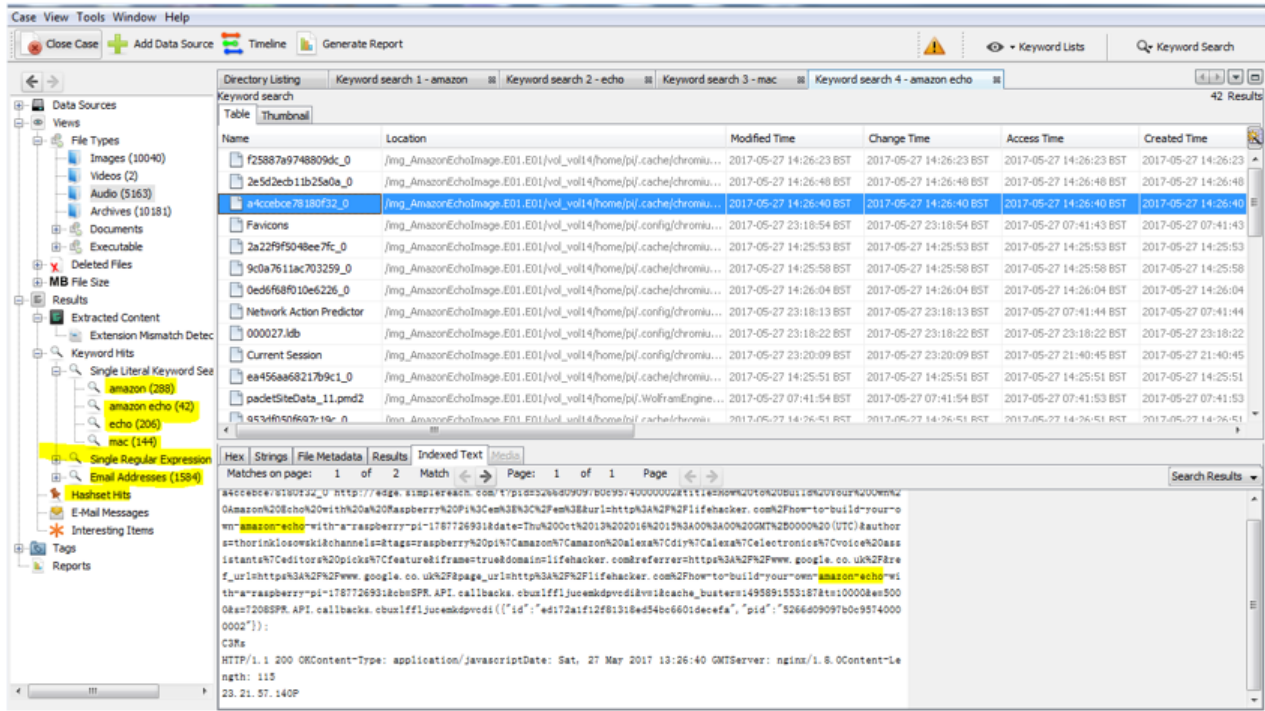
Fig. 11. Firmware analysis using Autopsy

The location data extracted from settings shows the whereabouts of the user. It also provides the geolocation data (e.g. address, postcode) that the user was searching for. However, during the analysis, the investigator also needs to check the history to get more context of the search request.

Amazon Echo is also capable of storing private conversations in the home, or other non-verbal indications that can identify who is present in the home (e.g. based on audible cues). However, in this stage we are only able to identify the recordings streamed by Echo from the user's home activated by the wake words. The text-transferred recordings are stored on both Amazon Alexa Server and Alexa companion app.

The device time zone is key to identifying data with an associated timestamp. In our further examination of the firmware, the device time zone can be used to validate the access/modify/creation time of files like .wavtemp. The wake word and language are also key to analyzing the history. Since the default wake word is Alexa, the investigator can also find the user-defined wake word, if any.

When examining the companion app and the firmware, we also located 5163 audio files (e.g. stop.mp3 and error.mp3), which can indicate the last operating time of the device (see Figure 11).

Figure 11 also shows the keyword search results. In this example, keywords such as amazon, echo and mac were used and 288, 206, and 144 results were found respectively. There are also 1584 potential email addresses located, which may contain the user accounts or potential passwords for logging to the AVS. In fact, we locate the login id with the corresponding password that can be used to login to the AVS, as shown in Figure 12. To further analyze the services that Echo provides, network forensics tools can be used to scan the open ports and potential services. In our examination, we use `Zenmap` with command {`nmap -T4 -A -v 192.168.0.10`} to scan the ports. We note that ports `80, 5200, 515, 427, 10001, 631, and 9100` are open. These open ports are very useful in analyzing the connection behaviors of Echo and can be used to trace the behaviors of the user. Using command {`nmap -sV -T4 -F 192.168.0.10`}, we found the services provided by Echo, including `http`, `svrloc`, `printer`, `ipp`, `jetdirect`, et al. Two additional services are also found using service fingerprints, and one of these two is shown in Figure 13.

The collective findings (from different devices, etc) are then pieced together.

5 DISCUSSION AND CONCLUSION

In IoT forensics investigation, evidence can be collected from IoT devices, internal network, applications, or outside cloud server, which makes it very challenging to identify evidences from the IoT systems. Basically, IoT forensics involves three levels: device level, network level, and cloud level. The heterogeneity in the IoT presented many challenges in terms of IoT forensics,

1) Identification of IoT devices and Data location, potential evidences may reside in the IoT devices, network infrastructures, and cloud server, it is very challenging but crucial to identify the devices and locations that potential evidences reside;

2) Data type and its lifespan, the data formant and type in IoT systems are very different for different type of devices, so it is very important to make sure the data will be returned to original format before performing forensic analysis. Specifically, it is very challenging to collect, process, analyse, and present data from live sources (*e.g.*, ram, cache, *etc.*) in an appropriate way that acceptable in law court;

3) Cloud level forensics, most IoT applications are delivered as services in cloud platform, in which the evidence can be distributed across different cloud servers that would be

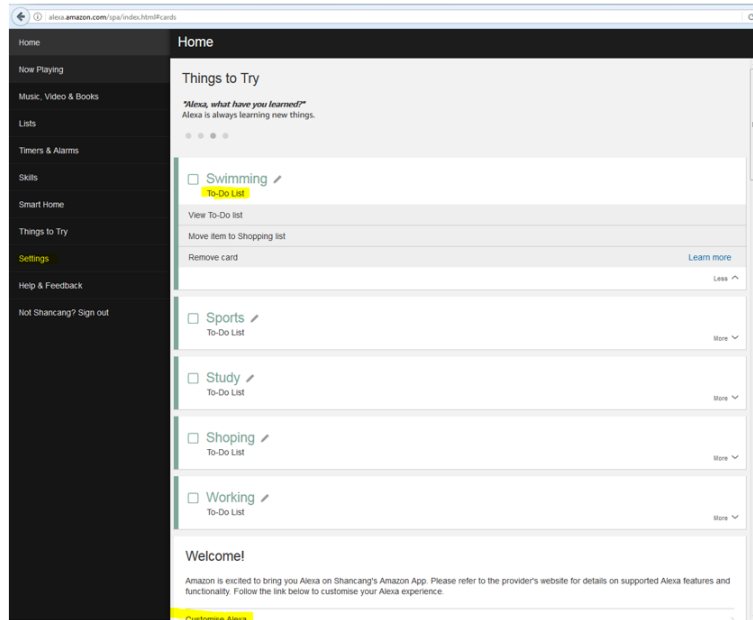Fig. 12. AVS website logged using the found login id and pwd from firmware

Fig. 13. Unrecognized services found in the images

outside of the investigator. Meanwhile, the dependence of forensically data on cloud level forensics require more powerful tools for forensic analysis.

Actually, there are still some extra challenges need to be addressed, such as security lack, device type, attack scenarios, *etc.* to develop methods/tools for IoT investigation.

IoT forensics will be increasingly important, as more devices around us are connected to the Internet or some form of networks (*e.g.* a private home or office network).

In this paper, we presented an IoT forensic analysis model and demonstrated how it can be used to guide the investigation of an Amazon Echo. There are a number of challenges that need to be addressed. For example, how do we address the large storage requirement associated with the search space? How do we keep pace with new IoT devices, in terms of our forensic capabilities and the potential for anti-forensic measures to be deployed?

Future research will also include evaluating our proposed model on a broader range of IoT devices and in cloud zone.

ACKNOWLEDGMENT

REFERENCES

[1] S. Li and L. Da Xu, "Security in enabling technologies," *Securing the Internet of Things*, p. 109, 2017.

[2] M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, "Low-cost security of iot sensor nodes with rakeness-based compressed sensing: Statistical and known-plaintext attacks," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 2, pp. 327–340, Feb 2018.

[3] J. Pawlick and Q. Zhu, "Strategic trust in cloud-enabled cyber-physical systems with an application to glucose control," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 12, pp. 2906–2919, Dec 2017.

[4] A. Nieto, R. Roman, and J. Lopez, "Digital witness: Safeguarding digital evidence by using secure architectures in personal devices," *IEEE Network*, vol. 30, no. 6, pp. 34–41, November 2016.

[5] H. Tsukayama. (2017) Bose headphones have been spying on customers, lawsuit claims. [Online]. Available: https://www.smh.com.au/technology/bose-20170420-gvo8pq.html

[6] J. Kastrenakes. (2017) Most smart tvs are tracking you vizio just got caught. [Online]. Available: https://www.theverge.com/2017/2/7/14527360/vizio-smart-tv-tracking-settlement-disable-settings

[7] Q. Do, B. Martini, and K.-K. R. Choo, "Cyber-physical systems information gathering: A smart home case study," *Computer Networks*, vol. 138, pp. 1–12, 2018.

[8] Q. Do, B. Martini, and K. K. R. Choo, "A data exfiltration and remote exploitation attack on consumer 3d printers," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 10, pp. 2174–2186, Oct 2016.

[9] K. P. Trommler. (2018) Know your enemy: What happens behind the scenes in a ddos attack. [Online]. Available: https://blog.paessler.com/types-of-ddos-attacks

[10] S. Zawoad and R. Hasan, "Faiot: Towards building a forensics aware eco system for the internet of things," in *2015 IEEE International Conference on Services Computing*, June 2015, pp. 279–284.

[11] S. Alabdulsalam, K. Schaefer, T. Kechadi, and N.-A. Le-Khac, "Internet of things forensics: Challenges and case study," *arXiv preprint arXiv:1801.10391*, 2018.

[12] A. MacDermott, T. Baker, and Q. Shi, "Iot forensics: Challenges for the ioa era," in *2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, Feb 2018, pp. 1–5.

[13] E. Oriwoh, D. Jazani, G. Epiphaniou, and P. Sant, "Internet of things forensics: Challenges and approaches," in *Collaborative Computing: Networking, Applications and Worksharing (Collaboratecom), 2013 9th International Conference Conference on*. IEEE, 2013, pp. 608–615.

[14] B. Copos, K. Levitt, M. Bishop, and J. Rowe, "Is anybody home? inferring activity from smart home network traffic," in *Security and Privacy Workshops (SPW), 2016 IEEE*. IEEE, 2016, pp. 245–251.

[15] O. Arias, J. Wurm, K. Hoang, and Y. Jin, "Privacy and security in internet of things and wearable devices," *IEEE Transactions on Multi-Scale Computing Systems*, vol. 1, no. 2, pp. 99–109, April 2015.

[16] W. W. Gibbs, "Build your own amazon echo - turn a pi into a voice controlled gadget," *IEEE Spectrum*, vol. 54, no. 5, pp. 20–21, May 2017.

[17] Q. Do, B. Martini, and K.-K. R. Choo, "Is the data on your wearable device secure? an android wear smartwatch case study," *Software: Practice and Experience*, vol. 47, no. 3, pp. 391–403, 2017.

[18] N.-A. Le-Khac, D. Jacobs, J. Nijhoff, K. Bertens, and K.-K. R. Choo, "Smart vehicle forensics: Challenges and case study," *Future Generation Computer Systems*, 2018.

[19] N.-A. L.-K. Arnoud Goudbeek, Kim-Kwang Raymond Choo, "A forensic investigation framework for smart home environment," in *In Proceedings of 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom 2018)*. IEEE, 2018, p. 1.

[20] L. Caviglione, S. Wendzel, and W. Mazurczyk, "The future of digital forensics: Challenges and the road ahead," *IEEE Security Privacy*, vol. 15, no. 6, pp. 12–17, November 2017.

[21] E. Oriwoh and P. Sant, "The forensics edge management system: A concept and design," in *2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing*, Dec 2013, pp. 544–550.

[22] C. Meffert, D. Clark, I. Baggili, and F. Breitinger, "Forensic state acquisition from internet of things (fsaiot): A general framework and practical approach for iot forensics through iot device state acquisition," in *Proceedings of the 12th International Conference on Availability, Reliability and Security*. ACM, 2017, p. 56.

[23] A. Nieto, R. Rios, and J. Lopez, "Iot-forensics meets privacy: towards cooperative digital investigations," *Sensors*, vol. 18, no. 2, p. 492, 2018.

[24] N. Chavez. (2017) Arkansas judge drops murder charge in amazon echo case. [Online]. Available: https://edition.cnn.com/2017/11/30/us/amazon-echo-arkansas-murder-case-dismissed/index.html

[25] H. Chung, J. Park, and S. Lee, "Digital forensic approaches for amazon alexa ecosystem," *Digital Investigation*, vol. 22, pp. S15 – S25, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1742287617301974

[26] C. S. Community. (2017) What is iot forensics and how is it different from digital forensics? [Online]. Available: https://securitycommunity.tcs.com/infosecsoapbox/articles/2018/02/27/what-iot-forensics-and-how-it-different-digital-forensics

[27] R. Hegarty, D. J. Lamb, and A. Attwood, "Digital evidence challenges in the internet of things." in *INC*, 2014, pp. 163–172.

[28] U. Salama. (2017) Investigating iot crime in the age of connected devices. [Online]. Available: https://securityintelligence.com/investigating-iot-crime-in-the-age-of-connected-devices/

[29] E. C. McLaughlin, "Suspect oks amazon to hand over echo recordings in murder case," 2017.

[30] R. A. S. B. W. Jansen, R. Ayers, and S. Brothers, "Guidelines on mobile device forensics," *NIST Special Publication*, pp. 800–101, 2014.

[31] V. L. Thing, K.-Y. Ng, and E.-C. Chang, "Live memory forensics of mobile phones," *digital investigation*, vol. 7, pp. S74–S82, 2010.

[32] S. Smalley and T. M. R2X, "The case for se android," *Linux Security Summit*, 2011.

[33] H. T. Al-Rayes, "Studying main differences between android & linux operating systems," *International Journal of Electrical & Computer Sciences IJECS-IJENS*, vol. 12, no. 05, 2012.

[34] J. T. Sylve, "Android memory capture and applications for security and privacy," 2011.

[35] D. Kirkpatrick. (2017) Gartner: Global wearables sales to grow 17 percent this year. [Online]. Available: https://www.marketingdive.com/news/gartner-global-wearables-sales-to-grow-17-this-year/503480/

[36] S. Li, L. Da Xu, and S. Zhao, "5g internet of things: A survey," *Journal of Industrial Information Integration*, 2018.

[37] B. Buchanan. (2017) The new digital investigator: Interrogating alexa. [Online]. Available: https://www.linkedin.com/pulse/new-digitial-investigator-interogating-alexa-buchanan-obe-phd-fbcs/

[38] Poonam Patil S and Rudrappa B Gujanatti, "AlexaPi on Amazon Ecosystem for Home Assistant Environment and IFTTT recipes," *International Research Journal of Engineering and Technology*, vol. 4, no. 8, pp. 1040–1042, 2017.

[39] "Geoffrey Bourne". (2016) I built my own amazon echo with a raspberry pi: Alexaberry. [Online]. Available: https://dzone.com/articles/i-built-my-own-amazon-echo-alexa-with-a-raspberry
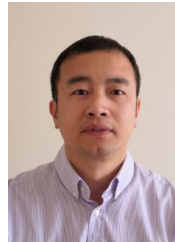
Shancang Li received the B.Sc. and M.Sc. degrees in mechanics engineering and the Ph.D. degree in computer science from Xi'an Jiaotong University, Xi'an, China, in 2001, 2004, and 2008, respectively.

He is currently a senior lecturer with department of computer science and creative technologies at University of the West of England, Bristol, UK. His current research interests include digital forensics for emerging technologies, cyber security, IoT security, data privacy-preserving, Internet of Things, Blockchain technology, and the lightweight cryptography in resource constrained devices. He has authored over 60 papers published in high profile journals and conferences. Dr. Li is the Associate Editor of IEEE Access and Journal of Industrial Information Integration. He is a member of the British Computer Society.

Kim-Kwang Raymond Choo received the Ph.D. in Information Security in 2006 from Queensland University of Technology, Australia. He currently holds the Cloud Technology Endowed Professorship at The University of Texas at San Antonio (UTSA), and has a courtesy appointment at the University of South Australia. In 2016, he was named the Cybersecurity Educator of the Year - APAC (Cybersecurity Excellence Awards are produced in cooperation with the Information Security Community on LinkedIn), and in 2015 he and his team won the Digital Forensics Research Challenge organized by Germanys University of Erlangen-Nuremberg. He is the recipient of the 2018 UTSA College of Business Col. Jean Piccione and Lt. Col. Philip Piccione Endowed Research Award for Tenured Faculty, ESORICS 2015 Best Paper Award, 2014 Highly Commended Award by the Australia New Zealand Policing Advisory Agency, Fulbright Scholarship in 2009, 2008 Australia Day Achievement Medallion, and British Computer Societys Wilkes Award in 2008. He is also a Fellow of the Australian Computer Society, and an IEEE Senior Member.

Qindong Sun received received his Ph.D. degree in School of Electronic and Information Engineering from the Xian Jiaotong University, China. He is currently the professor at the Department of Computer Science and Engineering of Xian University of Technology.

His research interests include network security, online social networks, digital forensics, cyber security, and internet of things (IoT). He is a member of China Computer Federation (CCF).

William J. Buchanan is a Professor in the School of Computing at Edinburgh Napier University, and a Fellow of the BCS and the IET. He was appointed an Officer of the Order of the British Empire (OBE) in the 2017 Birthday Honours for services to cyber security. In 2018 he received an Outstanding Contribution to Knowledge Exchange at the Scottish Knowledge Exchange Awards. One of his most recent achievements is the creation of a Blockchain Identify Lab and which is one of the first of its type in the world, and has significant industry funding.

Currently he leads the Centre for Distributed Computing, Networks, and Security at Edinburgh Napier University and The Cyber Academy (http://thecyberacademy.org). His main research focus is around information sharing, IoT, e-Health, threat analysis, cryptography, and triage within digital forensics. This has led to several World-wide patents, and in three highly successful spin-out companies: Zonefox (zonefox.com); Symphonic Software