Chapter 13
IoT Forensics

Sasa Mrdovic

Abstract This chapter provides an overview of research opportunities and issues in IoT forensics. It gives a quick introduction to forensics and digital forensics. Key specifics of IoT forensics are explained. Issues that arise from IoT related challenges in all phases of a forensic investigation are presented. Some opportunities that IoT brings to forensics are pointed out. An example of an IoT forensics case is provided. A detailed research overview is given, providing information on the main research directions with a brief overview of relevant papers. The chapter concludes with some ideas for future research.

13.1 Introduction

IoT, like any other system, needs a way to analyze things that happened within a system. When such analysis is performed for legal reasons it is called forensics. IoT brings many opportunities and issues in its forensics. Collection of forensic data from devices with very limited interfaces and capabilities for data storage and processing is challenging. On the other hand, aggregation of little data pieces from these devices can provide an unprecedented picture of events from various perspectives. That opens up a new chapter in digital forensics.

The future prevalence of connected things will provide an abundance of forensically relevant data. Our digitally connected lives leave traces that might lead to a golden age of forensics. Hard evidence will replace unreliable human recollection.

The opportunities of IoT forensics come with a price. The first issue that comes to mind is privacy. Fortunately, researchers are aware of this challenge and are working on addressing it, in general IoT and especially in IoT forensics. The focus of this chapter is mostly on the difficulties that a forensic investigator faces with IoT specific challenges.

S. Mrdovic (✉)
University of Sarajevo, Sarajevo, Bosnia and Herzegovina
e-mail: smrdovic@etf.unsa.ba

G. Avoine, J. Hernandez-Castro (eds.), *Security of Ubiquitous Computing Systems*,
https://doi.org/10.1007/978-3-030-10591-4_13

The chapter is organized in the following way. Section 13.2 provides a short introduction to general and then digital forensics, and ends with IoT forensics specifics. Open questions and hurdles that IoT forensics faces are presented in Sect. 13.3. Section 13.4 deals with opportunities that IoT forensics provide. An example of an IoT forensics case is presented in Sect. 13.5. An overview of IoT research with a focus on new approaches is the subject of Sect. 13.6. The last section presents conclusions and future research directions.

13.2 Forensics

Forensic science, usually called forensics, encompasses scientific methods used with the purpose of answering legal questions that generally arise in court cases and criminal investigations. One of the main activities in forensics is evidence collection and analysis. Evidence collection and handling has its procedures that should ensure that [286]:

- evidence is obtained legally, by court order or by order of an authorized institution or person;
- there is a chain of custody, which ensures that collected evidence is unaltered from the moment it was collected until the moment it is presented.

13.2.1 Digital Device Forensics

Digital forensics deals with evidence in a digital form. Digital, or sometimes called electronic, evidence is easy to change. Every access to a file on a digital device (PC, smart phone, IoT device) changes the file's last access time, and thus changes the file in a way that might constitute evidence. This is an example of a change that is neither malicious nor substantial but might be considered evidence tampering. An even bigger issue is the possibility of intentional malicious alterations. To ensure digital evidence integrity and its usability in court, procedures for electronic evidence and handling are in use [208].

The first step is the creation of a forensically correct evidence copy. The forensic copy is bit-by-bit identical to an original digital record. This must be done by tools developed for this purpose that are evaluated and tested to confirm their operation in accordance with a requirements specification [305]. Before and after copying bits, the tools create a cryptographic hash of the original and copied data to confirm data integrity during the copy making process. These tools also keep a log of all steps taken to ensure the existence of the mentioned chain of evidence. Further evidence analysis is performed on the copy, which ensures that the original evidence is unaltered and available in case it's needed for new analysis.

There is an ethical question of data privacy regarding digital evidence from any device that holds personal data. Digital devices are part of our everyday life. All of them, ranging from personal devices such as smart phones or personal computers to cloud and IoT devices, process and store huge amounts of data about personal users' private lives. Most of that data is not relevant in any particular case that triggered evidence collection from that digital device. Nevertheless, a forensic investigator frequently needs to examine a variety of files to be able to establish which ones are relevant to an investigation and which ones are not. Digital device examination is a significant invasion of privacy. For this reason, it is necessary to clearly define what kind of data and evidence an investigator should look for, to protect user privacy as much as possible. This issue in regard to IoT will be further addressed later.

The fact that digital evidence can be on a number of different devices in a variety of formats represents an additional challenge. These devices can be computers, smart phones, digital cameras, GPS devices, or any other device (thing) that stores data in a digital form (and it seems that it will eventually include everything). Each of these devices might store data on a different medium in a different way. Luckily there is some *de facto* standardization that usually makes it a little easier. Digital records on devices are created by different software that use different data formats. To read and understand a particular data format one needs specialized knowledge and tools. For this reason, evidence collection from special devices is often performed by a specialist in that area.

13.2.2 Other Digital Forensics

There are other sources of digital forensic data. Collecting data from those has some issues that are similar to IoT forensics. In reviewing them we can identify what differentiates IoT evidence collection.

All issues mentioned in the previous subsection relate to digital evidence that exist in the memory of a device. That memory is usually non-volatile. Volatile, working memory, like RAM, can contain forensically interesting data. The creation of an evidence copy of such memory must be done without powering down the device, this is known as live forensics. That procedure generally alters memory content and must be thoroughly documented.

Similar issues exist within network forensics. Some of the evidence can be collected from network devices, like routers, firewalls, etc., but most of it exists only in flight. That data can be captured only at the time it passes through a device processing it. There are devices and procedures for storing that network data. Nevertheless, it is impractical to capture and save all network data, due to its volume, however there are other issues such as the number and location of sniffing devices needed [174]. The question of privacy here is much larger as the network data might include a lot of information that is not related to the legal case in question [10]. Therefore, a narrow investigation focus is of utmost importance.

The more recent development in digital forensics is the need for cloud forensics. In its essence it should not differ much from device and network forensics, however it does. The cloud is built on a virtual infrastructure that is shared among many users of a particular cloud service provider. That presents a challenge for forensics because it is hard to locate, identify and separate data or virtual resources relevant for an investigation. In addition, there is no easy way to get access to the cloud infrastructure needed for creation of forensics copies. Data processing is decentralized and data cannot be collected without the cooperation of cloud service providers. Data decentralization might mean that parts of that data are stored in different jurisdictions where different authority can apply [483].

13.2.3 The Need for IoT Forensics

IoT forensics encompass these forensics: device, live, network and cloud. ‘Things’ might be devices with permanent storage with familiar file systems and file formats. Such ‘things’ can be treated as any other digital device. Unfortunately, ‘things’ might use proprietary file systems and formats. They might not even have permanent memory that holds user data and a limited power supply that severely limits duration or even prevents live forensics. ‘Things’ might have limited amount of RAM and transfer all their data immediately. That data can be transferred in an open standard or a proprietary closed format. Network data can be encrypted. IoT data is often processed in the cloud located in an unknown location that can be on the other side of the planet. All this makes IoT forensics different and more challenging than traditional digital forensics.

An IoT ecosystem, especially for forensic purposes, is divided into three areas: (IoT) device forensics, network forensics and cloud forensics [592]. Although all three of them are important, the focus of this chapter is on (IoT) device forensics. It is in line with the focus of this book on ubiquitous computing systems. The other two areas are more mature. Here they are briefly covered, to the extent required to understand IoT systems forensics.

One of the first papers on IoT forensics [455] created a list of differences to traditional digital forensics. A more recent paper [131], in an IEEE Security and Privacy issue devoted to digital forensics, extended that list. The differences found are the source of specific issues and opportunities in IoT forensics that are discussed in the next two sections.

13.3 Challenges in IoT Forensics

Issues specific to IoT forensics are systematized here based on the available literature. Initially, the general issues are presented followed by others as they occur in successive phases of a forensic investigation.

13.3.1 General Issues

There is a lack of a methodology and framework for IoT forensics. Even in digital forensics there is no single universally accepted methodology, but there are a few that are recognized and used by practitioners and researchers. IoT forensics is still in its infancy and relies on methodologies and frameworks from standard digital forensics that might not be fully adequate.

There is a lack of appropriate tools for IoT forensics. New software and hardware will be needed. A good overview of available forensics tools and their suitability for IoT is given in [559]. After a thorough analysis the authors concluded that existing traditional computer forensic tools are insufficient for cyber-crime investigations in IoT systems.

Since IoT is everywhere, it might be difficult to establish which jurisdiction a case might fall under as there will often be more than one involved. IoT systems can have devices in different jurisdictions as well as different cloud locations and providers. This is not too dissimilar to the Internet with its worldwide reach. IoT just expands the issue from the digital to the physical world.

13.3.2 Evidence Identification, Collection and Preservation

With IoT forensics the first thing to do is to identify the available sources of evidence. The investigator must establish which devices recorded relevant data. The question that needs to be answered is how IoT interacts with its surroundings. The investigator can then know which of the possible available sources to use. In addition, information on where and in which format data is saved must be obtained. Before collecting evidence, constraints to data collection (physical, proprietary standards, legal) should be checked.

Detecting the presence of IoT systems [258], and identification of IoT devices that can provide evidence in an investigation can also be challenging [264]. In addition, the device might contain data on different users not just the one(s) relevant to the investigation. Identification of data of a particular user is not an easy task.

A wide array of different devices makes it difficult to have a standardized approach to evidence collection. Data extraction is made difficult by the limited capabilities of devices, their various interfaces, and their storage formats. Data extraction without making changes to the device or data can potentially be difficult, which is an issue in forensics and can even be considered evidence tampering. On-board data storage is generally not accessible via traditional digital forensic methods [576]. There are limited methods to create forensic images of a given IoT device [152]. It can be difficult or even impossible to collect residual evidence from the device in a forensically sound manner [188, 405]. Encryption of data can make it difficult or impossible to collect evidence. Cumulative datasets may exist in multiple locations [576].

A crime scene in IoT involves physical things and the environment. It can be very difficult to preserve it in its entirety. IoT elements might be interacting autonomously. That can make it impossible to identify the boundaries of a crime scene and separate it from its surroundings [152].

There is a practical question if an IoT device needs to be kept as evidence or not. By its removal from the IoT environment there is an obvious loss of functionality.

13.3.3 Evidence Analysis and Correlation

The first issue for a digital forensics investigator when faced with an IoT system is how to analyze evidence from the physical world. IT knowledge might not be enough and expertise from related disciplines can be required [374].

The main issue at this stage of the investigation is the amount of data that an IoT system might produce. That amount can be overwhelming for an investigator [455] and the tools used [559]. The number of possible evidence sources in IoT is much higher than in standard digital forensics. Each source might produce a lot of data, for instance if it is a sensor that measures some physical property in small time intervals.

Since the evidence comes from a high number of heterogeneous sources it is more difficult to correlate. Creating a time-line is important in forensics. With a variety of devices with possibly unsynchronized clocks it can be very challenging to do [152]. All this is an issue in reconstructing events of interest. More evidence should mean better reconstruction but requires a lot of effort that might not be worth it [131].

The issue of privacy is most present in this phase. Aggregation and analysis enables pieces of evidence to be put together, and to establish someone's identity and actions. That is a good thing if the identity belongs to the person being investigated, however it is difficult to know in advance. Data collected from an IoT system might contain a lot of information on individuals not relevant to the investigation [440]. It is best to filter that data out at the time of collection but it is generally not possible due to time and resource limitations in that phase. Even the data on individuals relevant to the investigation might contain personal information that is not important. This issue also exists for standard digital forensics, however in IoT the data is collected continuously and indiscriminately from within the sensors' reach, and usually when individuals are involved this happens without their knowledge.

13.3.4 Presentation

Presenting forensic findings in a case involving IoT can be challenging. It is a new forensics and legal field. The courts are just learning to accept virtual evidence and this physical/virtual combination that IoT brings might be confusing.

There is also an additional question in this phase, as well as in the others, of whether a court will accept the methodology and tools used since they are not yet standardized.

There are several issues that are more relevant for forensic practitioners than researchers, but should be mentioned. How much court knowledge and understanding of IoT operations should be assumed? Should IoT devices be brought to court and an explanation provided on how they work before presenting evidence from them? Should an IT or an IoT expert present evidence? [374].

13.4 Opportunities of IoT Forensics

Fortunately, we do not only encounter challenges with IoT forensics. There are also opportunities. Some of them are presented in this section.

IoT brings new sources of evidence to general forensics. IoT records events from the physical environment, which were not recorded and stored before. They are now even stored as digital data. That enables much easier search, filtering, cross-relating, aggregation and other data operations that are helpful in turning data into evidence. IoT systems can contain contextual evidence collected without the individual who committed the crime being aware. This all happens automatically, without any user interaction as a side effect of the IoT operation [264].

IoT evidence, both for physical and digital forensics, is also harder to destroy [131]. It usually is not just one piece of evidence and it is generally stored in the cloud out of the reach of people who may want to delete it. As mentioned in the previous paragraph, usually suspects are not even aware of the evidence being collected. If that is the case they will not see the need and will not try to delete collected evidence.

IoT offers more evidence sources than standard digital forensics. Connected things provide an abundance of forensically relevant data. All devices that might collect, process, store or exchange data are interesting as possible sources of evidence. Even the smallest sensor that transmits a single value of measurement of a single physical quantity might be important. A composite picture of events can be constructed from all the data collected from the IoT systems. For example, the location of a suspect at a particular time can be established by correlating data from different IoT devices from various locations the suspect frequents. Wearable activity monitors can also help identify the approximate location of the suspect [592].

13.5 An Example of an IoT Forensics Case

To present how the above-mentioned IoT forensics challenges and opportunities can relate to a “real” case, a DFRWS IoT forensics Challenge will be used. The Digital Forensic Research Workshop (DFRWS) is a top forensics conference. It has a yearly forensics challenge, and in 2017 the focus was on IoT. The challenge is open to the public but it is particularly directed towards forensic researchers and practitioners. It is intended to motivate new approaches to IoT forensic analysis. The submitted solutions had to include source code openly available under a free software license with supporting documentation. Explanations of the procedure used to analyze the data needed for reaching conclusions were also required. There were four submissions. The winners were announced in May 2018. The challenge details and all submissions with explanations and tools used are available on a `github` repository for the challenge [302]. All people interested in practical aspects of IoT forensics are strongly advised to read the challenge and check out all proposed solutions. It is state of the art in this area at the time of writing. A lot can be learned from the solutions, explanations and tools.

The case scenario is simple. A woman has been murdered. Her husband has called an ambulance. The husband claims to have been at home at the time of the murder. Contestants had to analyze available artifacts for forensically interesting information and try to conclude who killed the woman.

An overview of general IoT forensics issues, analyzed in Sect. 13.3.1, as they relate to this case are given below.

- **Tools:** The fact that there is a challenge to develop new tools shows that existing tools, either free open source or proprietary commercial, are inadequate for IoT forensics. All submitted solutions used a combination of existing and tools specifically developed for this purpose.
- **Jurisdiction:** In this case there were no jurisdiction issues. Investigators had access to all data collected. In this case the husband provided credentials for cloud stored data. In reality, such credentials might be missing and a court warrant can be required to obtain cloud data that can potentially be in a different country.

Evidence identification, collection and preservation challenges, described in Sect. 13.3.2, which this case brings are presented next. In this particular case the investigators were provided with a list of digital devices found on the scene and the images or data from the devices including the cloud provider network traffic dump. A quick overview of issues police might have faced during relevant data collection is provided.

- **Identification of devices with relevant evidence:** There are some obvious devices that might contain relevant data such as the victim's mobile phone and the smart wristband she was wearing, as well as the husband's mobile phone. Since the husband claimed to have been watching TV using a Raspberry Pi as a smart device at the time of the murder, data from that Raspberry Pi is also of interest. There was a smart Amazon Echo speaker in the apartment that might have recorded something of interest for the investigation. Three sensors, a main sensor, a bedroom door sensor, and a motion sensor, connected to a Samsung SmartThings hub were found. A Google OnHub AP/Router provided Internet access and it had forensically interesting data. It was connected to the

Samsung SmartThings hub and an IPTime network switch. Finding all these devices required effort and knowledge given by the police officers who were on the scene. In this case they missed the smart power outlet.

- **Data extraction/image creation:** Getting relevant data from the devices in a forensically sound manner depends on the type of the device. Creating an image of a mobile phone can be challenging without credentials. Since a Raspberry Pi holds all its data on its SD memory card, imaging it is not difficult. IoT devices like the wristband and the sensors usually do not store any data. Data they generate can be found on the mobile phones used to control them, in the cloud of the smart service provider, or in a network traffic dump. Data collected from the Amazon Echo device can be obtained partly from the device and partly from the cloud. Access to the cloud data here was possible as the victim's husband provided the password. In some other cases, the password might not be available and cloud data would be more difficult to get. Network traffic is usually not logged on a permanent basis. In this case it was possible to obtain a diagnostic report from the OnHub AP/router in Google's protocol buffer specification format and a SmartHome network traffic dump for a period of one relevant hour.
- **Encryption:** It was not an issue in this case since the credentials for the devices and the accounts were available. In general, in the case of encrypted data on devices, or network dumps of encrypted traffic, collection of data in readable format might be impossible.
- **Multiple data locations:** Data was saved on multiple locations: devices and clouds.
- **Crime scene preservation:** Police arrived at the crime scene a short time after the relevant event. Data was collected in a timely manner and there was no need for additional crime scene preservation.

The practical issues of evidence analysis and correlation (Sect. 13.3.3) in this case are explained next.

- **Physical world data expertise:** There was no need for expertise on the data collected from the physical world. All events were simple (opening, closing, motion, steps) and were easy to interpret. That should be expected in current home automation, but in an industrial environment or a smart city this would have been different.
- **Amount of data:** The total amount of compressed data was over 6 GB. It was only a small household of two persons and a relatively short period of time. One can only imagine the data amount expected in a case involving an open space in a smart city.
- **Correlation/time-line:** The analyzed data was from six different devices and locations. For analysis, knowledge was required on how each of the devices and services, which were sources of data, worked and what the collected data meant. Significant effort was required to establish how the devices were connected and how all the data correlated. The time-line was a little easier to establish since all the devices had working and fairly synchronized clocks.

- **Privacy:** Analysis revealed a lot of personal data on the victim and her husband. The history of their phone usage, which included messages, application usage, locations visited, and health data collected from the health devices, among other information, was available to the investigators. Some of their physical conversations, the commands they gave to their smart devices, their movement through the home, and their TV viewing habits were also disclosed. Some data, such as the victim's phone messages and Echo-recorded conversations, were crucial for the investigation, but some, such as the husband's TV viewing habits, were not.
- **Effort needed:** The challenge was open for 3 months. Even after that period, none of the four teams that submitted solutions could be certain that their explanation of the events was completely correct. In total 18 highly skilled professionals worked for 3 months. It was a significant effort and as it was a murder case a human life was at stake. In real life, it is hard to imagine having so many experts available for an investigation. In the author's opinion, that fact might be the biggest practical problem of IoT forensics. Real forensic investigations are not like the CSI program we see on TV.

The challenges of how IoT forensics can present findings, as described in Sect. 13.3.4, are described below for this particular case. The details of the solutions that the competing teams provided were very technical. They were difficult to read and understand for non-experts. They were not written for the court, but for an expert panel. In reality, an additional effort would be needed to prepare the case presentation in front of a judge and jury. In this case it should have been possible to explain to the general public the sequence of events established after the forensic investigation. Technical details need to be kept to a minimum and should be presented as an appendix to the written report.

This case clearly shows the opportunities that IoT forensics can bring. Without IoT forensics the main, and probably only, evidence would have come from eyewitnesses. There might not have been any eyewitnesses, or they could have been hard to find. The case would have had to rely on memories that might have been incomplete, unreliable and subject to change.

IoT forensics enabled the collection of more reliable evidence. That evidence was used to reconstruct the sequence of events that led to the murder. The sensors provided data on door opening and movement. That data enabled investigators to establish the presence of a third person. The smart loudspeaker data contained parts of relevant conversation that confirmed the presence of a third person arguing with the victim. The victim's phone data showed unfriendly conversation between the victim and the coworker she seems to have been having an affair with. That enabled us to establish the identity of a suspect. Data from the smart TV confirmed that the husband was watching TV at the time of the attack on his wife.

13.6 Research Overview

Here we present and analyze the research directions in the literature that address the above-described issues and proposed solutions. IoT forensics research is in its early stages. It did not really start until 2013. It seems that researchers are just beginning to scratch the surface of this vast area. This opens up opportunities for young researchers to join in and offer fresh ideas. Most of the published papers expand on previous results in standard digital, network, cloud and mobile forensics. Several research directions can be identified. The following overview is organized into subsections, with papers that propose similar ideas grouped together and ordered by publication year. Some papers might belong to multiple subsections but are referred to in just the one that corresponds to its main contribution.

13.6.1 New Models and Frameworks

The biggest number of papers belongs to this group. It is understandable as models and frameworks need to define and provide some direction and standardization for researchers and professionals. Unfortunately, none of the proposed models and frameworks has been widely accepted and most of the proposals are still of theoretical nature. A brief overview of papers follows.

In addition to defining challenges and IoT differences, as mentioned in Sect. 13.2.3, [455] proposed some approaches to address challenges. Authors proposed a zone-based method for approaching IoT related investigations. They call them 1-2-3 zones. Zones loosely correspond to three areas of IoT forensics: device, network and cloud. Zone 1 is an internal network with all devices and software. Zone 2 covers hardware and software at the network border that provides communication services from outside networks. It can include a firewall and IDS. Zone 3 is everything outside of the network being investigated and includes cloud and ISP, among other things. Zones enable work in parallel or they focus on the one that is most urgent. Authors also propose a preparation phase for IoT forensics methodology. Future papers confirm the need for the phase where data collection devices are installed in advance. The paper proposes another important concept, Next Best Thing (NBT). It can be expected that in IoT some sources of evidence will not be available or reliable. The NBT model suggests that forensically interesting data can be acquired from devices that are either directly connected or somehow related to the object of forensic interest, as authors call it.

Another IoT digital forensic investigation model is presented in [470]. The model divides IoT into a number of zones, similar to [455]. It includes concepts for base device identification, a location finder represented by zones. The conceptual idea is on the right track, but the paper does not propose how the model could be implemented.

Yet another framework is proposed in [318]. That framework is created to comply with ISO/IEC 27043:2015, the international standard for incident investigation principles and processes. The authors hope that their approach to the standardization and creation of a framework will enable tool development. Their framework consists of distinct processes: proactive process, IoT forensics, and the reactive process. The proactive process is similar to the preparation phase in other models and frameworks. IoT forensics is the same as in [592]. The reactive process consists of initialization, acquisition and investigation which happens after an incident is identified in an IoT-based environment.

Harbawi and Varol [258] propose a theoretical framework that should improve the evidence acquisition model for IoT forensics. They address the problem of the identification of the main source of digital evidence in IoT. A Last-on-Scene (LoS) algorithm is proposed, consisting of seven steps for things-of-interest identification. It extends ideas from [455] and [470]. After things of interest are defined, a modified digital forensic procedure consisting of seven steps is proposed. The authors also propose an online management platform that manages and clusters IoT digital forensic cases, but the paper does not elaborate further than platform general specifications.

Zia et al. [602] propose adding application specific forensics to digital forensics model. They argue that to ensure the collection of evidence in the context of specific IoT applications it is important to have application-specific forensics in place. This application specific component feeds data into the digital forensics component of their model. It provides relevant data for the IoT application in question. That enables focused extraction of artifacts relevant to the investigation. The authors picked the top three most popular IoT applications at the time of writing: Smart Home, Wearables and Smart City. For each of these they defined items of forensic interest in a complete IoT system: device, network and cloud. With this approach the IoT forensic process should be focused on important data but is still holistic.

Privacy protection in IoT forensics is the focus of [440]. It proposes a Privacy-aware IoT forensics model (PRoFIT) that takes into consideration the privacy requirements established by the ISO/IEC 29100:2011 privacy framework. The model, similarly to others, relies on a preparatory phase. In this phase a piece of software may be installed to assist and advise the user about the information contained in the device according to privacy policies and forensic restrictions. Data collection is based on informed user consent. The logic is that enough IoT users will provide this consent and have the software installed. In that case at the time of the investigation a lot of data will be readily available. There might always be a need for court-ordered data collection but less than without the preparation step. The model presents privacy protection aspects through the rest of the investigation process. It includes asking the user's consent whenever there is a need in the investigation to share user data with someone who was not included in previous consents.

The same authors combined their work on PRoFIT [440] with Digital witness [442] to advance IoT forensics while providing user control of private data in [441]. Digital Witness is, exactly what its name suggests, a device which is able to collaborate in the management of electronic evidence. To stimulate a willingness among citizens to collaborate and allow their devices to be Digital Witnesses they need assurance about the protection of their personal information on such devices. The paper shows how it is possible and feasible to have a PRoFIT-compliant Digital Witness. The authors evaluate and confirm their approach in two cases: social malware and warehouse registration.

13.6.2 Preparation Step with Repository

Most of the proposed models and frameworks define the need for a preparation phase in IoT forensics. Such a phase has been suggested for other types of digital forensics. In IoT, due to the lack of logging and local data preservation, a type of pre-prepared local data repository of potential evidence seems to be the most warranted. The following papers offer some ideas on how this can be achieved.

The work in [456], from the same authors as [455], proposes a concept that introduces a device in between the local IoT network and the home firewall (Internet/Cloud) that provides security and forensic services. It is an end-user-managed solution which is unusual and has its pros and cons. The device provides standard security services like IDS/IPS, network monitoring, logging and threshold establishment. Once something happens that causes a crossing of a set threshold the forensic services are activated. They include data compression, parsing and differentiation, storage, time-line creation, alerting, preparation and presentation of results. It is an interesting idea but relies on the end user who might want to hide certain events. Authors also mentioned issues, common with all systems that monitor network traffic, when encryption, compression and steganography are used on the network data. In addition, a device that sits in the path of network traffic might become a bottleneck.

The FAIoT paper [592] formally defined IoT forensics and listed its challenges. It proposed a forensics-aware model for the IoT infrastructures (FAIoT). The model is supposed to help researchers focus on a specific research sub-problem of the IoT forensics problem domain. The FAIoT consists of three parts: a secure evidence preservation module, a secure provenance module, and access to evidence through an API. The authors propose a centralized trusted evidence repository as a new service available for all the IoT devices. Since the repository should handle very large datasets, the authors propose the use of the Hadoop Distributed File System (HDFS). To ensure a proper chain of custody by preserving the access history of that evidence, a provenance aware file system [433] for the repository is proposed. FAIoT could help with IoT forensics investigation but at this stage it is more a conceptual design than a practically usable system.

One more paper that proposes a preparatory phase in order to obtain evidence is [405]. It argues that collection of IoT devices states can enable an investigator to create a clear picture of the events that have occurred. The authors propose a centralized controller that can be connected to devices, controllers (hubs) and the cloud to acquire IoT states. This controller can only read the state and cannot change it. It logs states with time stamps in secure storage with hashes for integrity. The authors provide a proof of concept implementation of their controller using an open source IoT device controller, OpenHAB. It is connected to one device (IP camera), one controller (Insteon Hub) and a cloud account for a device (Nest thermostat). Analysis of the logged states enables reconstruction of scenarios of events in the physical world.

Wang et al. [572] propose a system that ensures data provenance. It ensures that it is possible to establish where a piece of data came from as well as the processes and methodology by which it was produced. It enables a holistic explanation of the system activities, including malicious behaviors. With data provenance, the sequences of activities in an IoT system can be connected with causal relationships. It replaces isolated logging and the analysis of individual devices. The authors implemented this idea with a centralized auditing system for a Samsung SmartThings platform called ProvThing. They showed that, through optimization, real-time system auditing is possible with minimal overhead.

13.6.3 Real-World Systems

The paper "Digital forensic approaches for Amazon Alexa ecosystem" [145] discusses practical issues when carrying out the forensic analysis of IoT systems present in many households. It proposes a combination of cloud-native forensics with forensics of companion devices, and this is called client-side forensics. Since Alexa is a cloud based assistant, most of its data is in the cloud. The authors used unofficial Alexa APIs to access its cloud data. Analysis of network traffic with the Alexa cloud-using proxy, enabled the authors to establish that the data are returned in the JSON format. This reveals issues with cloud forensics retrieving data that might not be available in its raw form but only through calls to predefined query functions, and these functions might not be documented. In this manner the authors were able to obtain some Alexa-native artifacts. Alexa is usually managed through a mobile application or the web. The authors apply forensics of mobile applications and web browsers to retrieve additional artifacts from the client. To automate this process of data collection, visualization and evaluation, the authors created CIFT (Cloud-based IoT Forensic Toolkit). The paper emphasizes the need for a holistic approach to data collection and analysis. The DFRWS Challenge, presented previously, had a practical scenario with more popular home IoT devices, including Alexa Echo, a smart speaker that is part of the Alexa Echo system.

In [494] the authors investigate what data can be collected in IoT attacks and what can be reconstructed from that data. They used a hub and sensors from "sen.se" as well as the Samsung hub, both for the home environment. After a 20-day period of operation, the collected data was analyzed. The authors explain the challenges they faced even with this small system. The paper shows how with a small number of simple sensors different attack scenarios can be identified, interpreted, data preserved and analyzed and presented in a way that is easy to understand. It clearly demonstrates the power and opportunities that come from IoT forensics.

13.7 Conclusion and Future Research Directions

IoT forensics is a new area open for research. There is already a need for practical solutions to questions that arise during investigations that include IoT. That need will help advance the research through practice.

The omnipresence of IoT makes it an ideal tool for evidence collection. As part of its normal operation IoT collects data from its surroundings. That data, from different IoT devices, can be correlated to create a very detailed reconstruction of events. Suspects can easily be unaware of recordings and usually cannot destroy the evidence.

Although this seems like a dream come true for surveillance agencies it can be perceived as “scary” for the common citizen. IoT forensics, with data correlation, can enable the emergence of personally identifiable information from, what seems like, irrelevant pieces of data. The issue of privacy in IoT is a very important one and needs to be adequately addressed.

With the opportunities it offers, IoT forensics has also some issues. New devices, new interfaces, new storage media, new file systems, new network protocols, dispersed cloud storage, unclear authority and jurisdiction are just some of those. The amount of data that needs to be preserved, stored and processed is huge. Even the presentation of the results can be challenging.

IoT forensics research so far has three main directions: the creation of new models and interfaces, the creation of systems with a pre-prepared repository for evidence, and forensics of real-world IoT systems. It seems to be at its beginning, so there are definitely many opportunities for further research.

From the author’s point of view two new technologies, blockchain and SDN, can play an important role in that future research. IoT’s distributed nature seems to be a good fit for blockchain which is built to, among other things, ensure integrity which is important for forensics. SDN can play a key role in relevant IoT traffic filtering which can be set up on an ad hoc basis.