FAIoT : Towards Building a Forensics Aware Eco System for the Internet of Things

Shams Zawoad and Ragib Hasan
{zawoad, ragib}@cis.uab.edu
University of Alabama at Birmingham
Birmingham, Alabama 35294-1170, USA

Abstract—The Internet of Things (IoT) involves numerous connected smart things with different technologies and communication standards. While IoT opens new opportunities in various fields, it introduces new challenges in the field of digital forensics investigations. The existing tools and procedures of digital forensics cannot meet the highly distributed and heterogeneous infrastructure of the IoT. Forensics investigators will face challenges while identifying necessary pieces of evidence from the IoT environment, and collecting and analyzing those evidence. In this article, we propose the first working definition of IoT forensics and systematically analyze the IoT forensics domain to explore the challenges and issues in this special branch of digital forensics. We propose a Forensics-aware IoT (FAIoT ) model for supporting reliable forensics investigations in the IoT environment.

Keywords—IoT Forensics, Forensic Investigation, IoT Security

I. INTRODUCTION

The pervasive nature of the Internet and the cost effective miniaturization of smart electronic devices introduces a new computing paradigm – the Internet of Things (IoT). The IoT is considered as the future evaluation of the Internet that realizes Machine-to-machine (M2M), Radio Frequency Identification (RFID), context-aware computing, wearable and ubiquitous computing [1]. The basic idea of the IoT is to allow autonomous and secure connection and exchange of data between real world devices and applications [2].

Everyday, hundreds of physical things get connected with the Internet to share local information to cyberspace. The US National Intelligence Council (NIC) estimates that by 2025 Internet nodes may reside in most of our surrounding things food packages, furniture, paper documents, and many more [3]. According to a report by Gartner, there will be 26 billion IoT devices in next five years [4]. International Data Corporation (IDC) forecasts that the IoT market will reach $3.04 trillion and there will be 30 billion connected things in 2020 [5]. These things can be varied in different attributes: processing and computation power, communication medium, dimension, etc. [6].

However, the rapid growth of IoT also brings some new challenges in terms of security. Since billions of things are interconnected to perform personal as well as business related activities, attackers may find IoT a very attractive target of attacks. A malicious individual can also launch attacks from the IoT environment. European Police Office (Europol) states that the world’s first death caused by the IoT is expected to occur before the end of 2014 [7]. An attacker can exploit the weakness of crucial health and safety equipment or the communication channel and trigger malicious instructions to jeopardize a patient’s life. To investigate such attacks, we need to execute digital forensics procedures in the IoT paradigm, which we refer as IoT Forensics.

Unfortunately, digital forensics in the age of IoT is challenging because the existing digital forensics tools and procedures do not fit with the IoT environment. The large number of IoT devices will generate massive amount of possible evidence, which will bring new challenges for all aspects of data management. Investigators will find it very challenging to collect evidence from the highly distributed IoT infrastructures. The wide variety of IoT devices will also raise problem in data analysis because of the heterogeneous formats of data. Reliability of the evidence can also be questionable since the attacker can tamper with the evidence resided in the IoT devices. On the other hand, the IoT can offer new opportunities to investigators. Since the IoT devices share local physical information, an investigator can use such information to establish facts about a criminal incident.

In this paper, we address the IoT forensics from two perspectives: first, executing digital forensics procedures in the IoT infrastructures, when the IoT is target of attacks or used to launch an attack; second determine new/unknown facts by utilizing the IoT infrastructure. Based on the characteristics of IoT and digital forensics procedures, we identify the challenges of executing each of the processes of digital forensics, where the evidence remains in the IoT infrastructures. We propose FAlIoT – a forensics-aware model for the IoT infrastructures to support both of the perspectives.

Contribution: The contributions of this work are as follows:

- To the best of the authors knowledge, this is the first work to formally define IoT forensics. The two different perspectives of IoT forensics introduced here can spawn future research in this area.
- We systematically analyze the challenges and opportunities for IoT forensics and propose a model for forensics-aware IoT – FAlIoT. Our analysis can help researchers to focus on specific research sub-problems of the IoT forensics problem domain.

Organization: The rest of the paper is organized as follows: Section II provides the background knowledge about digital
forensics, IoT, IoT forensics, and a hypothetical case study. In section III, we present the challenges of IoT forensics. Section IV presents the FAIoT model. Section V presents the related work and finally, we conclude in Section VI.

II. BACKGROUND

In this section, we first present a brief overview of digital forensics and the IoT. Next, we define IoT forensics and present a hypothetical case study of digital forensics involving the IoT environment.

A. Digital forensics

Before 2006, there had been no separate US Federal law for using electronically stored information (ESI) as evidence in civil cases. Federal Rules of Civil Procedure (FRCP) broadened the scope of evidence in the 2006 amendment and included ESI to be used in civil litigation [8]. FRCP defines the discoverable material and under this definition, data stored in hard disk, RAM, or Virtual Machine (VM) logs, all are discoverable material for the forensic investigation. According to a definition from NIST [9], digital forensic is “an applied science to identify an incident, collection, examination, and analysis of evidence data”. From the above working definitions, we note that digital forensics comprises four main processes:

- **Identification**: There are two main steps in identification: identification of an incident, and identification of the evidence, which will be required for successful investigation thereof, with potential correlation to other incident(s).
- **Collection**: In the collection process, an investigator extracts digital evidence from various media (e.g., hard disk, cell phone, e-mail, and many other types of data). The investigator also preserves the integrity of the evidence.
- **Organization**: There are two main steps in the organization process: examination and analysis of the digital evidence. In the examination phase, an investigator extracts and inspects the data and its characteristics. In the analysis phase, he or she interprets and correlates the available data to come to a conclusion, which can serve to prove or disprove civil, administrative, or criminal allegations (when interpreted legally).
- **Presentation**: In this process, an investigator makes an organized report to state his or her findings about the case. This report should be appropriate for presentation to the competent court or proceedings.

In digital forensics, maintaining the integrity of the information and strict chain of custody for the data is mandatory. Several other researchers define computer forensic as the procedure of examining computer system to determine potential legal evidence [10], [11].

B. Internet of Things (IoT)

The idea of Internet of Things (IoT) was first proposed by MIT’s Auto-ID centre [12]. The International Telecommunication Union’s (ITU) Internet report 2005 formally proposed the Internet of Things [13]. According to the report, we are heading towards the age of ubiquitous network society, in which networks and networked devices are omnipresent. All of our surrounding things will be interconnection through the Internet of things for data interchange; These things include personal computers, laptops, tablets, smart phones, insulin pump, tires, refrigerator, television, air cooler/heater, and many more. By 2020, there will be 10 connected IoT devices for every person of the world and 40 to 80 billion IoT devices in total [14]. Most of the IoT devices embed different sensors and actuators that can sense, perform computation, take intelligent decisions and transmit useful collected information over the Internet.

The Internet of Things involves many heterogeneous technologies. Among them, RFID (radio frequency identification) and wireless sensor technology are most mature. There is a wide range of applications for IoT, such as home appliance control, health care management, automotive services, inventory management, and many more. In general, the IoT devices capture data from the physical environment through different sensors and sends the data to the cloud for intelligent decision making or for other data processing tasks.

C. IoT forensics

We define IoT forensics as an especial branch of digital forensics, where the identification, collection, organization, and presentation processes deal with the IoT infrastructures to establish the facts about a criminal incident.

We identify IoT forensics as a combination of three digital forensics schemes: device level forensics, network forensics, and cloud forensics, which are illustrated in the Figure 1.

- **Device level forensics**: An investigator may need to collect data from the local memory of the IoT devices. When a crucial piece of evidence needs to be collected from the IoT devices, it involves the device level forensics.
- **Network forensics**: The source of different attacks can be identified from network logs. Therefore, network logs

![Fig. 1: IoT Forensics](image)
can be very crucial to condemn or exonerate a suspect. IoT infrastructures includes different forms of networks, such as Body Area Network (BAN), Personal Area Network (PAN), Home/Hospital Area Networks (HAN), Local Area Networks (LAN) and Wide Area Networks (WAN). Important piece of evidence can be collected from any of these networks.

**Cloud forensics:** One of the most important roles in the IoT forensics domain will be the cloud forensics. Since most of the IoT devices have low storage and computational capability, data generated from the IoT devices and IoT networks are stored and processed in the cloud. This is because cloud solutions offer various benefits including convenience, large capacity, scalability, and on-demand accessibility.

**Hypothetical Case Study.** Alice is suffering from high blood sugar and she always wears a blood sugar monitor device. At her home, there are other smart devices, such as heating system, television, refrigerator, intelligent medicine dispenser, car, etc. All of these devices are connected with the Internet and are controllable from Alice’s mobile device. Alice also works in a hospital, where there are thousands of health care related IoT devices and the hospital allows its employees to connect their smart devices with the hospital’s network. Mallory creates an intelligent malware to collect data from the smart health care devices. First, it infects Alice’s smart refrigerator, gets connected with the Alice’s blood sugar monitor through the shared network, and finally, infects the blood sugar monitor. Later, when Alice goes to the hospital for work, the malware searches for other devices which shares the same network as the blood sugar monitor. In this way, Mallory is able to infect hundreds of smart health care devices located in the hospital and steals confidential electronic medical records (EMR). When the data breach gets identified, Bob, a forensics investigator is assigned to investigate the case. The number and variety of IoT devices available at the hospital will make Bob’s investigation very challenging. Bob needs to execute device level forensics for all the available devices. Later, he needs to investigate network logs for all the devices to identify the source of infection. This will not only includes the smart health care devices but also the smart mobile device that the health care professionals generally bring everyday.

### III. Challenges in IoT Forensics

The traditional tools and technologies of digital forensics are not designed to completely handle the IoT infrastructure. In this section, we identify the challenges in each of the steps of digital forensics, while dealing with the IoT environment.

#### A. Identification

The billions of IoT devices will generate massive amount of data. When the amount of possible evidence is very large, it is difficult to identify the important pieces of evidence that can be used to determine the facts about a criminal incident. For example, there are thousands of IoT devices in a hospital and only one of the devices may get comprised and leak confidential electronic medical record (EMR). However, the number of logs generated from the thousands of IoT devices of the hospital can be very large, and finding evidence to identify the compromised device can be treated as finding a needle in the haystack.

### B. Collection

After identifying the evidence, investigators need to collect the evidence to analyze and find the facts. Any errors that have been occurred in the collection phase will propagate to the evidence organization and reporting phase, which will eventually affect the whole investigation process. Hence, this is one of the most crucial steps of forensic procedure.

Some of the factors that make the data acquisition process in IoT forensic harder than traditional computer forensics are discussed below:

The Internet of Things involves numerous connected nodes with different technologies and communication standards. An investigator may find it very challenging to collect evidence from this highly distributed infrastructures. For a hospital environment, where there are thousands of IoT devices, it may not be even possible to collect evidence from all the IoT devices of the hospital in a short time period.

Because of the storage limitation of IoT devices, most of the data generated by IoT devices are stored in the cloud. Since clouds will be one of the main sources of evidence for IoT forensics, some of the problems of collecting evidence from clouds apply for the IoT forensics, such as physical inaccessibility to clouds. The established digital forensic procedures and tools assume that we have physical access to the computing resources, e.g., hard disk, network router, etc. However, in cloud forensics, the situation is different. Sometimes, we do not even know where the data is located as it is distributed among many hosts in multiple data centers. A number of researchers address this issue in their work [15], [16]

Since confidential information, such as EMR can be leaked from the IoT, the collection process may suffer from legal issues, especially the privacy and data protection laws. These laws can vary depending on the jurisdiction. It may happen that a forensic investigator is in one jurisdiction and the data reside in another jurisdiction, where the privacy laws of these two jurisdictions are not in harmony. Therefor, organizations need to carefully consider the legal ramifications of where they store and process data to ensure that they remain in compliance with the regulations that they face [17].

### C. Organization

Till now, there is no widely accepted protocol or standard for IoT. Vendors are using their proprietary protocols for things-to-things communication. The wide varieties of structure of the data generated by the IoT devices make the examination and analysis phase challenging.

Analyzing logs from different sources plays a vital role in digital forensic investigation. Process logs, network logs, and application logs are really useful to identify various malicious activities and the users behind those malicious activities. A standard format of logs can make the data organization phase smooth. Organizing logs collected from different sources (such
as multiple IoT devices) is challenging, as there are no standard formats for logs across different systems. Some of the logs may not even provide crucial information for forensic purpose, e.g., who, when, where, and why some incident was executed [18]. We could correlate logs collected from different IoT devices and identify crucial information if there was a standard format for logs.

Within the IoT, we expect to see an explosion of data because of the increased number of interconnected devices that will be communicating and exchanging information across the IoT information highway. Organizing such a big dataset to identify the facts about a criminal incident can be challenging. For a very large dataset of evidence, manual review and decision-making cannot work. Often time, it requires special data mining techniques and tools to identify the facts [19], [20].

D. Presentation

The final step of digital forensic investigation is presentation, where an investigator accumulates his/her findings and presents to the court as the evidence of a case. Challenges also lie in this step of IoT forensics. Providing the evidence in front of the jury for traditional computer forensics is relatively easy compared to the complexity of managing IoT data. Jury members possibly have basic knowledge of personal computers or at most privately owned local storage. But the technicalities behind identifying, filtering, analyzing data from highly distributed and heterogeneous IoT environment can be far too complex for them to understand.

IV. FORENSICS-AWARE IOT AND OPPORTUNITIES

In this section, we first present a model for forensic-aware IoT (FAIoT). Then, we present new opportunities available for the digital forensics investigators from the FAIoT.

A. The FAIoT Model

Since the IoT infrastructure is highly distributed and there is no standardization among the devices, we propose a centralized trusted evidence repository in the FAIoT to ease the process of evidence collection and analysis. The evidence repository will also apply the secure logging scheme [21] to ensure the reliability of the evidence. We can consider this as a new service available for all the IoT devices. The devices just need to register this secure evidence repository service. Figure 2 illustrates our proposed model.

Secure Evidence Preservation Module: This module will constantly monitor all the registered IoT devices and store evidence securely in the evidence repository. Evidence can be network logs, registry logs, sensor readings, etc. While preserving the data, this module can take care of segregating the data according to the IoT devices and its owner. In this way, multiple users’ data will not be co-mingled. This module will also preserve the confidentiality of the data from malicious cloud employee by using public-private key based encryption, so that only investigators can view the data.

Such repository needs to handle very large dataset. Hence, we propose to use Hadoop Distributed File System (HDFS) for the evidence repository. Hadoop\(^1\) is an open source implementation of Google’s proprietary MapReduce framework. The HDFS is the file system component of Hadoop, which is designed to store very large data sets reliably, and to stream those data sets at high bandwidth to user applications [22]. Therefore, HDFS will perform better than the traditional database management systems when retrieving a small piece of information from a big dataset of possible evidence.

Secure Provenance Module: This module ensures the proper chain of custody of the evidence by preserving the access history of the evidence. Using provenance aware file system (PASS) [23], the evidence repository can produce the provenance record for evidence usage. However, as all the evidence and the access history are under the control of secure evidence repository provider, they can always tamper with the provenance record. Moreover, from the provenance data in cloud, an attacker can learn confidential information about the data stored in cloud. To protect provenance information from these types of attack, we need a secure provenance scheme. This module will apply secure provenance chaining [24] to

\(^1\)http://hadoop.apache.org/
preserve the integrity of the provenance record.

**Access to Evidence Through API:** We propose to provide secure read-only APIs to law enforcement agencies. Only the investigators and the court will have access to these APIs. They can collect the preserved evidence and the provenance information by calling these APIs. To implement this feature, the secure evidence repository provider need additional web server, which will communicate with the previously described modules to collect the requested data by an API call. The web server provides Representational State Transfer (REST) based API using the synchronized data, and the provenance record as resources. To retrieve these evidence, GET operations can be used on the resources. Caller of a REST service can pass different parameters to retrieve his desired result.

**B. Opportunities**

The availability of FAIoT can also brings new opportunities for digital forensics investigators. Since IoT share physical information to the virtual world, this can give new information to the investigator and can help to identify new insights for a criminal activity.

From the various on-board sensors and logs of these IoT devices, we can identify important information of the surroundings of the devices that can help us to determine facts about a criminal incident. For example, by correlating data of different IoT devices, we can determine the location of a suspect at a particular time. From the data of the IoT devices of the suspect’s home or office building, we can determine whether the suspect was at either of these places. From the wearable activity monitors’ data, we can also identify the approximate location of the suspect. An ideal solution to handle large amount of sensor data will quickly identify crucial information of a criminal case.

**V. RELATED WORK**

Security of the IoT infrastructure has been addressed by researchers. Oren et al. showed that a smart TV could be compromised using a cheap antenna and through broadcasting messages, as it relies on an insecure Hybrid Broadcast-Broadband Television Standard (HbbTV) [25]. Researchers from IOActive Labs [26] presented a mechanism that can be used to attack traffic control systems. Magnetic sensors used in the streets (to collect and disseminate data) could be compromised using professional transmitters, or antennas from a couple of miles away, as there are few security protocols in place.

In [27], researchers presented an access delegation method with security considerations based on Capability-based Context Aware Access Control (CCAAC) model intended for federated machineto-machine communication or IoT networks. By using the identity and capability based access control approach along with the contextual information and secure federated IoT, the proposed model provides scalability and flexibility as well as secure authority delegation for highly distributed system. Ning et al. proposed a cyber-physical-social based security architecture to deal with information, physical, and management security perspectives, and presents how the architectural abstractions support IoT model [28]. In [29], researchers examined the digital evidences collected from mobile devices and proposed a scheme of correlating mobile phone’s location with the contacts based on the incoming and outgoing calls.

Researchers propose several solutions to ease the process of digital forensics in the cloud. To make the network, process, and access logs available to customers, Bark et al. proposed to expose read-only APIs by CSPs [30]. By using these APIs, customers can gather valuable information and can provide this to investigators. Recently, Dykstra et al. implemented FROST [31], a forensic data collection tool for OpenStack. Using FROST, cloud users/investigators can acquire an image of the virtual disks associated with any of the user’s virtual machines, and validate the integrity of those images with cryptographic checksums. It is also possible to collect logs of all API requests made to CSP and OpenStack firewall logs for users’ VMs. To get necessary logs from IaaS cloud model and to preserve the integrity and confidentiality of the logs, Zawoad et al. proposed Secure Logging-as-a-Service (SecLaaS) [21]. By using this service, investigators can collect various important logs, e.g., network, process, registry, and application logs. SecLaaS can also detect any alteration of logs by a malicious CSP, or a malicious forensic investigator. Thorpe et al. developed a log auditor by using the ‘happened before’ relation [32] in the cloud environment [33]. Delport et al. focused on isolating an instance to mitigate the multi-tenancy issue in cloud forensics [34].

The closes work related to our work is presented in [35]. In this paper, the researchers presented the design of the Forensics Edge Management System (FEMS), a system that autonomously provides security and forensic services within the home Internet of Things (IoT) or smart home context. FEMS focus on the user-defined solution to manage the IoT-based evidence. While providing users’ control, the proposed system also provides malicious users the opportunity of tampering with the evidence.

**VI. CONCLUSION**

The rapid increase of IoT devices creates new attack surfaces. Hence, there is a growing need on providing forensics supports in the IoT environment. However, the IoT infrastructure, large number of IoT devices, and wide variety of the IoT devices impose new challenges for the digital forensics investigators. In this article, we define the term IoT forensics and identify the challenges of executing reliable forensics in the IoT domain. We also propose a conceptual model for executing digital forensics in the IoT infrastructure. Solving all the challenges of IoT forensics can open the opportunity of identifying many new insights that were not possible before.

**ACKNOWLEDGMENT**

This research was supported by the National Science Foundation CAREER Award CNS-1351038, a Google Faculty Research Award, and the Department of Homeland Security Grant FA8750- 12-2-0254.
REFERENCES


