Journal of Universal Computer Science, vol. 25, no. 9 (2019), 1199-1218
submitted: 14/1/19, accepted: 31/7/19, appeared: 28/9/19 © J.UCS

Digital Investigation of IoT Devices in the Criminal Scene

François Bouchaud
(IRCGN - Forensic science laboratory, Pontoise, France
francois.bouchaud@gendarmerie.interieur.gouv.fr)

Gilles Grimaud, Thomas Vantroys
(Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL, Lille, France
{gilles.grimaud,thomas.vantroys}@univ-lille.fr)

Pierrick Buret
(C3N - National cyber-crime unit, Pontoise, France
pierrick.buret@gendarmerie.interieur.gouv.fr)

Abstract: The Internet of Things (IoT) is everywhere around us. Smart communicating objects are offering the digitalization of lives. They create new opportunities within criminal investigations. In recent years, the scientific community sought to develop a common digital framework and methodology adapted to IoT-based infrastructure. However, the difficulty in exploiting the IoT lies in the heterogeneous nature of the devices, the lack of standards and the complex architecture. Although digital forensics are considered and adopted in IoT investigations, this work only focuses on the collection. The identification phase is quite unexplored. It addresses the challenges of locating hidden devices and finding the best evidence to be collected. The matter of facts is the traditional method of digital forensics does not fully fit the IoT environment. Furthermore, the investigator can no longer consider a connected object as a single device, but as an interconnected whole one, anchored in a cross-disciplinary environment. This paper presents the methodology for identifying and classifying connected objects in search of the best evidence to be collected. It offers techniques for detecting and locating the appropriate equipment. Based on frequency mapping and interactions, it transfers the concept of "fingerprinting" into the field of crime scene. It focuses on the technical and data criteria to successfully select the relevant IoT devices. It gives a general classification as well as the limits of such an approach. It shows the collection of digital evidence by focusing on pertinent information from the Internet of Things.

Key Words: Internet of Things, Digital Forensics Model, IoT Forensics, IoT Investigations, Evidence[1] acquisition
Category: B.0, D.0, D.2, D.4.4, E.0, H.1.1, H.3.1, H.4.3

1 Introduction

With the development of connectivity between many objects, the spreading of new communication protocols like LoRa and Sigfox and the cost-effective miniaturization of smart electronic devices, the Internet of Things is taking place within our daily lives. Uses are diversifying and affect almost all areas. The rapid growth of IoT brings security and forensic challenges. Billions of things interconnected with private and business data are attractive targets for attacks.

[1] In this paper, the term "evidence" must be understood in legal and forensic sense.

In the meanwhile, this hyper-connected universe makes available to all the digitalization of life. It creates some opportunities for criminal investigations. Traditional methods of recording events sometimes prove to be wrong because they depend on people susceptible to errors, prejudices according to the period and the context. In many cases, the element collected by eyewitnesses has not been a major source of truth and contributes to more than 70% of the mistakes, which are then contradicted by scientific evidence, like DNA. The IoT increases the volume, variety, and veracity of direct evidence of human activity. Thus, IoT can bring much more than localization and identity through a dense network of sensors, but it also obtain the necessary information on actions performed by individuals or by a device on a crime scene. This information may be decisive in a court of law or may guide the investigation, for example, to date an event, to find a murderer or to confuse testimonies. Several real cases integrating connected objects have recently made headlines. A very interesting example is the case of Anthony Aiello in San José. It is the match between the victim's FitBit data and the information out of the home automation system that has set light on the murder of a woman.

With the heterogeneity of connected objects and the lack of standards, this paper develops a methodology to identify and classify IoT devices. The purpose is to help the investigators and prioritize the collection of information within an IoT infrastructure. In this article, we will use the home device called "*Sen.se Mother*" as a way to illustrate our proposal. This paper is organized as follows: **Section 2** describes the concept of the Internet of Things and the link between evidence and connected objects. **Section 3** focuses on taking into account the IoT infrastructure on a crime scene. **Section 4** presents a methodology to answer this problem. **Section 5** explains how an investigator can prioritize these devices based on the amount of collectable data and **Section 6** shows collection methods.

2 Background

In this section, we present a taxonomy of the Internet of Things. We suggest matching the connected objects with the elements sought during an investigation of the crime scene.

2.1 A definition of the IoT

[Dorsemaine et al. 2015] defines the IoT as a "group of infrastructure interconnecting connected objects and allowing their management, their data mining and the access to data they generate". Each object is uniquely identifiable, carrying out a specific role and interoperate within the network (Figure 1). The interconnection of objects brings advanced services.

To illustrate this architecture, we rely on the use case derived from the IoT smart home called "*Sen.se Mother*" [Sen.se 2018]. It consists of a local pickup point called "*Mother*" and tags called "*Cookies*". Tags can be used in the home to

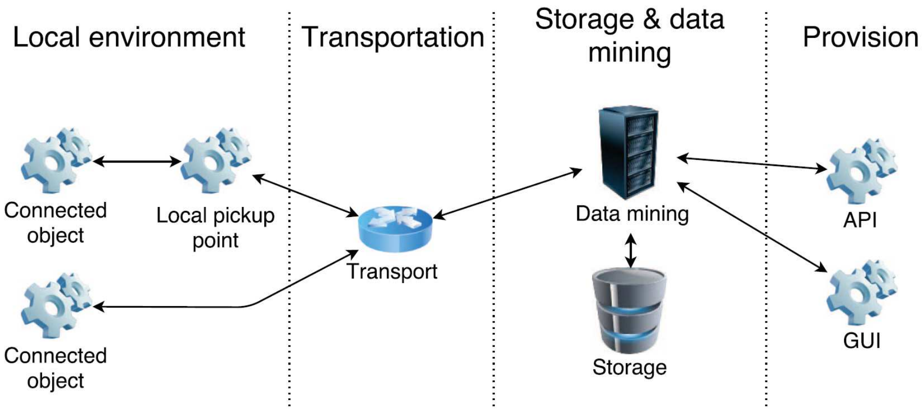

Figure 1: Architecture related to the IoT, described in [Dorsemaine et al. 2015]

turn a simple object into a connected object. The "*Mother*" collects the sensor feedback information to be sent to the Internet or triggers scheduled actions. Furthermore, this gateway connects peripherals from other platforms such as Philip Hue. Data are stored online on the platform Open.Sen.se and can be displayed through mobile phone applications. A set of Web APIs is also available to create applications that exploit data collected by "*Cookies*". Based on the CC430F6137 System on a Chip (SoC), the "*Cookie*" provides an integrated thermometer sensor and a three-axis accelerometer.

2.2 IoT classification based on data characteristics

[Rahman et al. 2017] explore a new approach of IoT devices based on the types of data collected. The paper presents a forensic data classification defined with two IoT devices "*Sen.se Mother*" and "*Samsung Hub*". This concept is widely used in IoT crime scenarios and attempts to identify grounds of evidence.

Connected objects collect physical data from their environment, through different sensors. They send them to the cloud for data processing tasks. We identify three categories of evidence searched by investigators, and each of them is associated with sensors families and type of data (Figure 2).

The **presence** of a person on a crime scene can be revealed by **temperature** variations or by **detection** sensors. This event is **dated** and **located** precisely. The radio link can be considered as an event sensor. A communication links the timestamps to a specific operation. A change in the positioning of an object leads to the restructuring of the network architecture. Measurement variations inform the investigators of an event about when and how long this event has happened. In the example of a smart home, "*Cookies*" can be used for multiple purposes, like tracking the movement of an object or a person, with the **time** and the **duration** of the movement. The signature of the **movement** is captured, analysed and recognized by the "*Mother*" in order to carry out a specific action.

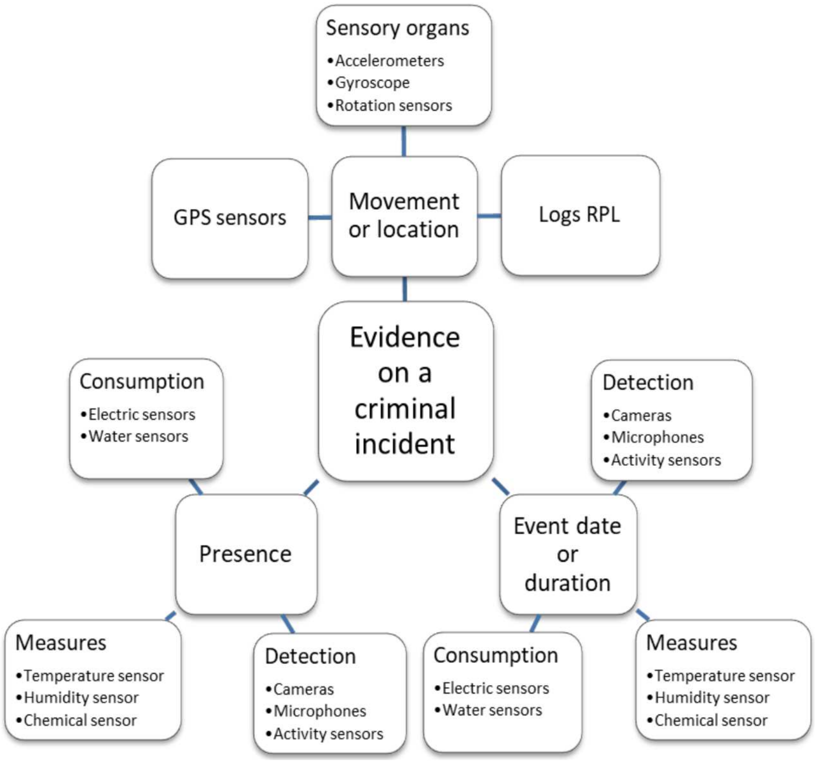

Figure 2: IoT categorization matching evidence to devices

The tag can also be used to control the bedroom's door openings or the **presence** of a person in a bed.

3 IoT investigations

In the context of digital forensics[2], investigators are well trained to identify the medium (phone or computer) to collect and to analyse it. Because IoT devices come in many forms, the identification process becomes more complex. Devices are heterogeneous, not always identifiable by lack of standards and even hidden. So, the investigator can miss the presence of an information system at the crime scene. The detection, localization and identification arise as serious issues. Moreover, the connected object is often selected according to its mechanical aspects and visible interests. The system data or the logs should attract the investigator's attention.

[2] According to a definition from the National Institute of Standards and Technology (NIST) [Kent et al. 2006], digital forensic is "an applied science to identify an incident, collection, examination, and analysis of data while preserving the integrity of the information and maintaining a strict chain of custody for the data".

3.1 Related works

The search for evidence in an IoT environment is an extension of the work and traditional uses of digital forensics. Although the field of IoT in digital forensic research is relatively new, there are already many interesting works. [Hegarty et al. 2014] discuss the fundamental challenges for the forensic with IOT devices. [Oriwoh et al. 2013] question how digital forensics in an IoT environment is moving away from the traditional approach. It set forth a forensic model based on zones approach to divide the crime scene into zones.

[Perumal et al. 2015] take up this concept of dividing the environment into zones. He gives direction on how to conduct the analysis and investigation of this concept. [Copos et al. 2016] collect network data from an IoT device from Smart Home Network Traffic. [Zawoad et al 2015] explain a Forensics-Aware IoT model (FAIoT). That is a kind of centralized trusted evidence repository after directly collecting evidence from the cloud, accessible to investigation by a specific interface. [Rahman et al. 2017] show a new approach based on attack scenarios in order to identify general sources of evidence. Thus, the related works follows the tracks of forensic collection and analysis by updating the methodology of digital forensics or by studying its architecture. However, the process of identifying and searching IoT devices, visible and hidden in the local environment, is poorly studied. This fundamental step is the basis of the investigation to find evidence for the court of justice.

3.2 Problem of identification

According to the definition of the National Institute of Standards and Technology (NIST), the identification phase includes two steps of analysis: an incident and evidence. The first step is important for successful investigation and a potential correlation with other similar events. The second step is not always immediately accessible. Indeed, the crime scene is dealt of visible and hidden devices, which broadcast the data in a timely manner. Moreover, connected objects are not always visually identifiable. The variety of devices, their roles in the IoT infrastructure, and their reliance on the network, especially for the data storage management policy, make it difficult to find evidence. Interesting data can be scattered over the infrastructure. Thus, the localization of the devices and the shape of the evidence are studied to allow an efficient recovery of these.

Given a huge amount of data to predict possible evidence, investigators must find effective processes and tools to select them. They cannot collect and analyse everything. Identification is made at the beginning of the investigation and during the collection phase. It must be adapted according to the elements found. In the collection process, an investigator identifies, labels, records and acquires live data from the sources of relevant data. Several difficulties are perceptible, particularly because of the network topology related to diversity of protocols and the presence of volatile memories. Many connected objects in IoT infrastructure use Real-Time Operating Systems (RTOS) which do not persist data. The acquisition of evidence may not be reliable due to the transient nature of IoT ad hoc network connectivity. The study of connected objects on the crime scene allows us to apply the existing solutions of digital forensics to perform the acquisition of volatile memories. Several case studies are presented in the last section.

4 IoT identification approach in a crime scene

In this section, we suggest a methodology to discover and identify devices on the crime scene before collection.

4.1 Digital footprint concept

The forensic investigation is based on two main principles: the concept of transfer with the principle of Locard [Locard 1920] and the concept of individuality from traces left at a crime scene, according to Kirk's vision [Kirk 1974]. The exchange principle of Locard stipulates that the contact between two items induces an exchange. The author shows something on the crime scene and comes up with something else, both of which can be used as forensic evidence. Since physical contact may not always be mandatory, this may extend Locard's principle to virtual interactions. However this exchange may be transient and easily lost. Kirk developed the concept of individuality from traces. Individuality implies that each entity, person or object, cannot be identical to itself. It is therefore unique. Two objects, natural or artificial, cannot be exactly the same. This vision has limits on the question of falsification and alteration of the information by external acts. The fingerprint is one of the most used characteristics to identify a person. It does not change over time and its numerous variants enable it to be sorted. The fingerprint of a digital system is its electromagnetic signature defined by its own hardware and software characteristics. Specifically, it can be made up of a set of technical characteristics such as the transmission frequency range, modulation, the shape of the signal, response time, addressing, signal power, etc. The imperfect transfer of unique features has an impact on the identification of a device. This loss of information may be related to environmental or temporal constraints, communication difficulties, external contamination, etc.

4.2 Definition of the investigation methodology

When an investigator arrives at the crime scene, he carries out several methodical operations. First, he protects the whole scene against external pollution. This approach is simple and obvious in the case of physical exchange such as fingerprints, firing residues or blood. It is more difficult in the digital context. The investigation process breaks down into four steps (Figure 3). Each step is applied to a "*Sen.se Mother*" device.

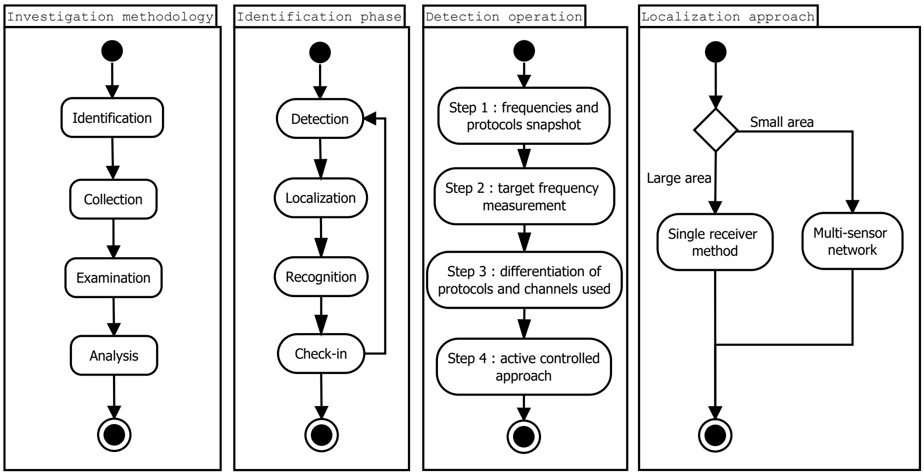

Figure 3: Investigation methodology for the IoT

In the detection phase, investigator studies the behaviour of an environment. He captures the electronic signature of space, defined by signal characteristics. This information gives an electromagnetic baseline and shows active devices. This first approach can be supplemented by external actions on the idle devices. Nevertheless, an impact study must be carried out beforehand. In our example, "*Cookies*" communicate with the "*Mother*" at 915.0 MHz on one channel, whose the Occupied BandWidth (OBW) is equal to 362.74 kHz. This first step makes it possible to deduce the presence of a connected device on a crime scene without knowing its position, its nature and the number of devices present, knowing that the "*Mother*" can accept a maximum of 24 "*Cookies*".

The localization of IoT devices is based on signals captured at several points. This approach consists of defining the origin of the signal by trilateration and by comparing the captured information. Studying the signal strength at several points allows the investigator to assess the position.

The recognition process is conducted from the visible information on the target and the frequency information captured. Correspondence tables allow you to perform reconciliations. In our example, the "*Cookies*" FCCID is 2ABGNCOO001 and for the "*Mother*" 2ABGNMOM001. On the other hand, each "*Cookie*" is individually identified by a unique name consisting of eight letters and numbers, such as 6FEB205D or AB61092C. The study of this individual signature allows the investigator to know the number of connected objects present at the crime scene.

Check-in for a list of potential evidence is performed. The investigator studies the interconnection devices. He determines the role of each element: actuators, sensors, nodes of the local network, gateways, etc. A dependency graph approach is performed. In your example, locally "*Mother*" communicates with "*Cookies*" over the wireless network and uses an Ethernet connection with the Internet. The identification of "*Mother*" and network settings on the crime scene allows investigators to retrieve information from the IoT platform and identify the user interface, including associated devices, such as a mobile phone.

4.3 Detection operation

The detection of active devices is done in four steps. This phase consists in determining the frequencies used and associating device families according to the study of the protocols. Before any detection operation, the investigator must remove all communication equipment that he is carrying: mobile phones, work computers, control equipment, etc. Indeed, these devices can disrupt the crime scene. At the same time, the investigators retrieve the story of the equipment used by the first responders. This information is exploited by step 3.

The **first step** is to detect known frequencies and protocols. This operation is performed passively. The investigator obtains a first frequency map of the active objects. This digital signature is a unique fingerprint at a given time, similar to a snapshot. The frequency sweep is done at several levels, methodically. Thus, he performs a first series of measures around the crime scene to freeze the general atmosphere. With global measurements, the technician identifies the external signals. He also studies the impact of measuring instruments on the crime scene. This observation is performed throughout the technical operations. Then, in **step 2**, he calculates at several points the radio signal on a target frequency. This is done over a long enough period to obtain relevant information. So, with these signatures, he detects the presence of objects on a crime scene. This approach is based on an understanding and comparison of technical characteristics of different protocols. The differentiation of the protocols is done in **step 3**. The channels used by the target devices are studied one by one. The influence of the radio environment is taken into account, as well as the behaviour of communication equipment such as variable speed channels, fixed channels, low or high power, etc. The perception process depends on the sensitivity of the detection method that is used and the ability to differentiate the trace from the background. However, the map is incomplete for several reasons. Only active devices are detected. The measure is limited in time, partial or unachievable. Therefore, we offer other solutions to detect passive devices or a group of devices based on traffic generation. Three solutions can be considered: generating wake-up order, adding a connected object in the network or jamming the frequency. A brief interruption in the power of a node may cause a new synchronization of network objects. However, this operation destroys the data stored in the volatile memories and modifies the scene of the crime. **Step 4** consists of active actions mastered at the crime scene. These technical operations are performed after the retention of the information contained in the devices observed in the previous steps. To illustrate the concept of "wake-up order", we can cite the Simple Service Discovery Protocol (SSDP) used in IoT devices. A thorough study of the communication protocols must be carried out in order to understand the principle of data synchronization. The ultimate goal is to be able to "force" pairing. Add, jam or disconnect a device in an IoT infrastructure can generate radio traffic between devices. Periodically, devices send "Hello messages" to detect other members of the network. If there is no response, the communication links are cut off or the devices choose to reorganize. To illustrate this concept, we use Routing Protocol for Low-Power and Lossy Networks (RPL). Routing is based on the proximity of objects for the most efficient communication possible. Any modification of a device in the network can have an impact on the organization of the network. An interdependence link exists between objects. Jamming can also be used to force some protocols to use specific channels.

Detection operations are made more complex by the presence of new standard protocols. Only the type and flags of this communication are heard. Many of them use spectral scattering, high-speed frequency switching, modulation of the first signal sent in response, cryptographic communication without the possibility of distinguishing two devices.

4.4 Localization approach

Localization is one way to distinguish devices when protocols do not allow it. This phase aims to map, differentiate and classify the target equipment. For each device, a spatial reference system is calculated.

The study of the Received Signal Strength (RSS) and its phase allows the investigator to locate a device. This measurement operation is performed according to several points. A limit of this method resides by the disturbance of the signals from other objects. The strength, shape, and signal disruption depend on the device in a given environment. Moreover, the power of the broadcast signal depends on the frequency and therefore in its ability to pass through solid bodies. Rebound or attenuation phenomena are observable. So, good localization requires identification of the study area. The type and size of the materials of the environment are taken into account by a topological study with a laser or sonar technology. The aggregation of data collected is a complex, long and ingenious operation. There are interesting works about indoor localization concept. For example, [Lassabe et al. 2009] investigate about Wi-Fi indoor localization to detect and locate the mobile phone with the accuracy of the order of less than one metre. [Altini et al. 2010] and [Höcker et al. 2010] present a localization approach for Bluetooth devices, based on multiple neural networks. [Blumenthal et al. 2007] study ZigBee-based sensor networks and study a localization approach for these ZigBee devices.

The results are usually not precise to be directly exploitable. Temperature, humidity and building configuration at a given moment influence the measurements. Furthermore, in an indoor environment, the devices can be located on several floors. The investigator must perform an "empirical" research. Depending on the device found and its role in the infrastructure, the investigator must find devices related to the latter on the same network. We study two techniques to take into account the entire crime scene and its singularities: a single receiver method and a multi-sensor network approach with a set of remote sensors.

The single receiver method involves scanning the entire crime scene with a multi-antenna sensor high-speed. The localization of an object is performed by comparing information collected during the displacement of the sensor. The study of the active capacities of the target device by the realization of a ping offers a refinement of the measurements. This type of equipment is used on drones for wide or unreachable outdoor surfaces. To improve the efficiency, this measurement method can be coupled with information from operator networks. These reduce the study area.

The multi-sensor network consists of taking measurements from fixed and moving points. This system can integrate sensors with accelerometers, gyroscopes and external capabilities such as Global Navigation Satellite System (GNSS) solutions. The obtained result is a contextualized measure in space and time. The integration of the temperature sensor brings significant added value, especially for the study of wave propagation and thermal signatures. This type of equipment is used in a small area. The calculating devices are positioned at least three ends. To improve accuracy, a recording device can be placed in the centre of the study area.

It is essential to observe and put an end to any interference, either by localization or by countermeasures. Moreover, radio-frequency data transmission on a crime scene disturbs the observed environment, even for the calibration of passive listening devices. In order to avoid an alteration of the data, the operations are performed in promiscuous mode. Only a spectral analysis proving the absence of signals over a sufficient time makes it possible to deploy sensors on unused frequencies. There are many electronic attacks on the localization technologies. However, the operations are conducted on digital investigation of a crime scene. The check-in phase may allow the investigator to note this absence.

4.5 Recognition process

4.5.1 Protocol recognition

With protocol information, we refine technical characteristics as product signature and signal structure. Bluetooth mesh networking, Wi-Fi and ZigBee use Media Access Control (MAC). This addressing mechanism is used to identify the manufacturer and the device. The idea is to compare this identification information with the elements of the traffic generated. However, some technical constraints can be observed. Bluetooth equipment, for energy saving problematic, is paused automatically. There is no continuous broadcast. That's why, the frequency scan should be calculated over a long period of time. For Wi-Fi communications, to collect the MAC, we must be near the Wi-Fi network because of the hidden node problem. The header that contains the MAC isn't encrypted. This approach requires the recovery of correspondence tables between MAC and devices. The adversary can modify MAC. However, this case is quite rare during a judicial intervention on a crime scene and corresponds to a specific type of delinquency. Prior to any investigation, the investigator conducts work on the environment and the profile of the suspects. Depending on the results of its analysis, specialized units are requested. In the general case where this type of situation is handled, the information collected must be referenced with the physical data extracted from the crime scene. In case of doubt, the log and system analysis phase will confirm this intention to change address. The capture of MAC is considered in some countries as a violation of privacy if this information comes from objects located outside the crime scene. So, it may be necessary to locate the devices to be able to prove their presence in the study area.

4.5.2 Physical recognition

In some cases, it is difficult to recover the device credentials. The electromagnetic signature can allow this operation. This information characterizes the device in a unique way. This physical identification can be used on the crime scene to discriminate devices with the same protocol. The response time, the shape of the signal also contributes to this distinction. A device signature database can be constituted by this information. The electromagnetic signature can also give information about the manufacturer and country of manufacture of the device. To illustrate this concept, we can cite the case of the radio-frequency identification (RFID) and its magnetic electronic signature ([Romero et al. 2009] and [Tedjini et al. 2012]). The identification of the device also involves the development of a universal knowledge base. This database contains technical information about the various IoT devices. Depending on the country, the device transmitting waves contains a Federal Communications Commission Identification (FCCID). This information refers to specifications and technical characteristics.

4.6 Check-in

The recovered data is cross-referenced to the known data. The identification cycle is rehearsed until everything is identified. Searching for a primary node during detection can help investigators trace back to different branches of the infrastructure. With this information, they develop a relationship diagram between the devices. The role of the devices in the infrastructure is defined, so that interdependencies and localization of data storage are understood.

Once the identification phase is complete, the investigator is confronted with a huge amount of potential evidence. In order to prioritize the collection process, it is necessary to define an effective methodology to select the best things that will provide the best evidences.

5 Selection procedure for relevant objects

5.1 Data properties

We define four main criteria to guide the collection: the relevance, the accessibility, the position and the type of data. To illustrate these different criteria, we rely on the use case "*Sen.se Mother*" [Sen.se 2018].

Data relevance develops along three axes: the relationship to the event, time and space. The proximity to the event plays a role in the choice of the data to be recovered and therefore the objects to be collected. The data has a validity date limited in time. Indeed, it can disappear or lose its relevance according to its rank. This criterion is also related to the role of devices in the infrastructure. The devices can be active or passive with respect to an event. The greater the spatial, temporal and relational proximity, the more the data collected is relevant and accurate to the event. This information contextualizes the data to be collected. For example, the opening and closing of a door can be determinate using motion detection sensors. However, a simple slamming of a neighbouring door may trigger a capture event. Similarly, a sensor located somewhere in the house where the communication with the base "*Sen.se Mother*" is not optimized, a disconnection and a reconnection of the sensor can be observed periodically generating the sending of messages. Indeed, the application associated with the detection a presence: "is the cookie connected to the "*Mother*" ?" Thus, if "*Cookies*" are disconnected several times over long periods, the "*Mother*" interprets this event as departures and returns.

Data accessibility is also an interesting criterion to take into account. For each selected device, it is more or less difficult to access the data. The data may be protected by encryption-type protection systems or code obfuscation. However, conventional methods used in digital forensic can be applied to data stored in Human-Computer Interaction (HCI). Recovering data on "*Cookies*" can only be performed by hardware access, which requires advanced engineering work. The gateway "*Sen.se Mother*" has Ethernet access.

Data position focuses on where the data is stored. It can be stored on the sensor or transmitted over the network. Indeed, all devices don't have memory or are limited in capacity. So, the captured data can be on the crime scene or in the IoT cloud. The position of the data has an impact on the technical acts to be performed by the investigator on the crime scene. The data can be stored in volatile or persistent memory. To access data on the cloud, the investigator needs to ask the operators of the platform. He therefore needs to know the local devices associated with the data collected. "*Cookies*" can store about 10 days of data in case of connection loss to the base

"*Sen.se Mother*". As soon as the base is accessible again, the sensors transmit the data. The "*Sen.se mother*" retrieves them only to transmit them to a dedicated platform. The cloud processes the received data, interprets them, analyses them continuously to be able to provide services. Collected data is accessible by a web interface or a dedicated application on a mobile phone. Investigators can extract offline sensor data, extract event logs to analyse network activity, and examine the Application Programming Interface (API). From this local information, they can request access to " textit Sen.se" platforms.

Data type is broken down into three characteristics : direct, transformed or interpreted by humans. Direct data is a raw data collected by a sensor. The data is modified and contextualized by the device according to the defined parameters such as a notion of the threshold. This data can lead to a human interpretation. It comes from the log analysis and an observation of how the direct data has evolved over time reflecting an operational state of the object. For example, the presence of a person can be deduced from a recorded event resulting from an action requiring human manipulation in the room. Direct and transformed data will be more easily exploitable than interpreted data. This indirect data is based on a thorough knowledge of the devices and their nominal operation. So interpretation mobilizes a lot of ways to retrieve information and knowledge. "*Cookies*" collect raw data as the temperature or the detection of presence. These data are easily exploitable and identifiable for investigators and inform on an abnormal change in real time. For energy-saving reasons, the transmitted information is asynchronous. The connection logs found on the device "*Sen.se Mother*" can be used to date the transmitted information. This information must be studied with the actors of the communication and their roles. The connection logs make sense because those data are linked to a measurement sensor event. So this interpreted data mobilizes a lot of knowledge and requires deductive approach.

5.2 Weighting of Device

With the data properties outlined in the previous section, we define the relevance of information collection in the IoT infrastructure. This classification is developed in connection with the technical principle: the performance cost of operations and the impact on the quality of the results. The idea is to select the best evidence at the crime scene. The four columns of the table 1 are the four data criteria. The row is made up of the four sections of the IoT infrastructure: the sensors, the gateway, the cloud and the HCI. Criterion 1 corresponds to the strongest weight and 4 to the weakest. These different elements are broken down according to the following four notions:

Productivity is the effort we have to do to get the data;

Human cost is the execution time of the operation;

Engineering cost is the financial aspect of the operation;

Alteration is the impact of the operations performed on the devices.

5.3 Discussion

In this section, we explain our weighting. They have been defined by digital forensic experts and verified with real cases.

		Data relevance	Data accessibility	Data position	Data type	Total	#
Productivity	Sensor	1	4	4	3	12	2
	Gateway - Node	4	1	3	4	12	2
	IoT platform	2	3	1	2	8	1
	HCI (API - GUI)	3	2	2	1	8	1
Human cost	Sensor	4	4	3	4	15	3
	Gateway - Node	3	1	1	3	8	1
	IoT platform	1	2	4	1	8	1
	HCI (API - GUI)	2	3	2	2	9	2
Engineering cost	Sensor	4	4	3	1	12	3
	Gateway - Node	3	2	1	4	10	2
	IoT platform	1	1	4	3	9	1
	HCI (API - GUI)	2	3	2	2	9	1
Alteration	Sensor	4	4	3	1	12	3
	Gateway - Node	3	3	2	4	12	3
	IoT platform	1	1	4	3	9	2
	HCI (API - GUI)	2	2	1	2	7	1

Table 1: Device categorization based on data properties and technical principles

5.3.1 Productivity topic

To illustrate this weighting, take the case of measuring the ambient temperature with the "*Sen.se Mother*". The measurement is done locally by the "*Cookie*". Data collected from the environment is raw information that is not always readable at first because it is not transformed. Data accessibility requires complex operations, such as chip-off. The "*Mother*" only contributes to the transmission of temperature data to the IoT platform. Connection logs follow the event. The gateway is not very secure and has many open ports for easy access to data stored. HCI offers centralized formatted and interpreted information. However, this information has been selected and depends on the choices of the application designer. The temperature application returns the measured data as a graphical rendering.Data accessibility depends on the container like mobile phone or

Web access by authentication. These media contain advanced security, difficult to bypass. The cloud contains the manufacturer's database. It provides complete centralization of information and an overview of the infrastructure. Collection requires the intervention of a third person. It is conditioned by the way of selecting the right information. So data accessibility depends on the operator.

5.3.2 Human and Engineering cost topic

Pluralities of connected objects are present at the crime scene. Each device is unique and involves significant research and development costs for extracting and formatting information. This element is reinforced by the question of the proprietary formats of the collected data. The CC430F6137 which is in the "*Cookie*" is a micro-controller with 32 KB of in-system programmable flash memory and 4 KB of RAM. This component has a 128-bit Advanced Encryption Standard (AES) accelerator to secure transmitted data to prevent it from being intercepted. For gateways, the main cost is related to the interpretation of the collected information. Listening to ports allows us to determine open ports and services used to dump memory. The "*Mother*" has an Ethernet connection to exchange information with the IoT Infrastructure. It uses TCP ports 123, 443, 6514, 8482 and UDP 53. Memory is also accessible by chip-off because the information is not encrypted. It uses a Linux kernel in which connection logs and infrastructure configuration are kept. The financial and time costs resulting from requests from platform operators are partially controlled as they result from a contract. The requests transmitted to the operators must be precise. They require to know the technical information relating to the gateway in the form of MAC or model number. The main cost for the HCI results from the extraction of the application data. However, this approach relies on controlled processes within the framework of digital forensics. The application contained in the phone can be studied with conventional digital forensic tools.

5.3.3 Alteration topic

During the collection and retrieval phases of useful information, objects and data stored in memory can be corrupted. Hardware methods, such as chip-off or Boundary Scan (Test Access Port, TAP), used to retrieve component information are very physically destructive but guarantee extreme data extraction. The software approach can modify the nominal operation of the device and is likely to generate a write on the systems. Sensors are usually addressed using a hardware approach. For example, the "*Cookie*" has 4 KB of RAM. This RAM buffer is used when data transmission is not possible. This device is able to keep about 10 days of data in memory. Turning off the power causes loss of information. If the investigator is unable to retrieve the data, the connected objects must be kept alive and placed in a Faraday cage for later extraction at the laboratory. The software approach is used for the IoT platform. HCI and gateways combine the two approaches.

5.4 General information

From defined weights, we elaborate a device classification of IoT infrastructure (Table 2). The main result is that collecting evidence is easy from HCI that administers the IoT devices, or from the cloud operator that collects data. It is more difficult to get the same results when dealing with the sensors directly.

	Total by data	#
Sensor	51 *(12+15+12+12)*	4
Gateway - Node	42 *(12+8+10+12)*	3
IoT platform	34 *(8+8+9+9)*	2
HCI (API - GUI)	33 *(8+9+9+7)*	1

Table 2: Device classification of the IoT infrastructure

HCI is the best performing part. Because of its interface position, it converges the direct data of all connected IoT devices. The data returned to this interface is interpreted. Thus, they are easily exploitable by the investigators during the analysis. However, access to data depends on the type of interface. Thus, collection operations can be technically complex and time consuming, resulting in a risk of data corruption and significant costs in research and development especially on the application layer. This interface contains only data that the designer wants displayed.

The cloud also offers interesting performance through its work as a hub and data storage. It contains additional information about the HCI interface, including usage data for maintenance purposes. However, the data are deported to the crime scene and require the intervention of a third party to the investigation: the operator of the cloud. Thus, the manner of collecting the information is not controlled by the investigator. Moreover, beforehand, a precise knowledge of the elements to extract is necessary to locate the correct information to recover. This knowledge depends on the elements collected locally. Thus, the associated constraints are the time and distance factors.

The gateway mainly contains indirect information in the form of log files. The exploitation of these elements requires a good understanding of the network and the devices.

The sensors are the interface between the crime scene and the IoT infrastructure. They are the direct witnesses of events. This acquisition device contains raw data from the crime scene. However, the object has only the local information that is specific to it. Collection operations depend on the characteristics of the object and its access interfaces. So, in some cases, only an electronic approach is practicable. It can be technically complex and time consuming, resulting in a risk of data corruption and significant costs in research and development. Moreover, this information is not always readable and accessible. Some objects are equipped with an external storage card. In this case, the collecting operation is simple. Some sensor-type devices also have a gateway role in storing information.

Prioritizing devices within the IoT infrastructure during data collection may be questionable according to the context of the incident. For example, investigators find a sensor on the crime scene that is not synchronized with the local network. These devices may contain information still present in the sensor. The IoT architecture is sometimes conceived in accordance with the fog or edge computing principles. The computing and analysis resources are distributed locally between the source and the cloud. Only processed data is transmitted to the cloud. In this case, the raw information is hosted by the fog node like gateways or by local devices. In some cases, the IoT infrastructure does not contain a cloud. HCI can connect directly to the gateway or to the sensors to display information. The classification must be comprehended in a general way. Some of the wide variety of IoT settings moderates our generic classification. Thus, we have determined the objects to be recovered in priority. It still needs determining how to recover them.

5.5 Use case

In this section, the use of wellness data from the users as critical evidence in court is examined. A body of a person is found in the bed of his house. When the police discover the crime scene, the causes and the context of death are not defined. No material or visual element makes it possible to choose the hypothesis of a homicide or a suicide and to reconstruct the chronology of the events. The study of the body and the room allows investigators to obtain more or less approximate study tracks. They freeze the crime scene and begin an identification phase. Frequency mapping is used to infer the presence of a Bluetooth 4.0 Low Energy connected devices. The device, located under the pillow of the victim, is found with the multi-sensor network approach. Visually, the device is identified by the brand symbol, FCCID 2ADIOB501, CMIITID 2016DP2433, model name B501 and unique ID 17302162. Connected object is a sleep sensor called Sleepace Dot.The investigator defines a dependency between the Sleepace Dot and a mobile phone. The study of the crime scene makes it possible to determine the associated smartphone. This element will be confirmed during the device analysis. At the local level, the investigator collects the Sleepace point and the smartphone by disconnecting it from the network. The gateway role is performed by the smartphone.

During his analysis, the investigator wants to understand what happened and establish a schedule of events. The study of the Wellness Coach app on the mobile phone provides personal information about the user, his activities and the operation of the connected object. The investigator retrieves the timestamps of the triggering of the sensor and the measurements of the movements of a body on the bed for several months. However, only manually synchronized information is available on the application. Thus, the latest days have not been found on the phone application. These early works contain information on the context and habits of the victim. From the identifier of the device, a judicial request is sent to the company to obtain the data stored in the cloud. The study of the 'gateway' elements is carried out during the analysis of the telephone, especially to study the connection between the sensor and the smartphone and the synchronization events. However, these operations require advanced technical knowledge in interpreting the results. So, the sleep sensor in the absence of synchronization with the application plays a decisive role in the resolution of the criminal investigation. Indeed, it keeps the information recorded over a period of seven days. From the extracted data, the interviewer is able to reconstruct the chronology of events, dating the activity of the victim and its interactions with the environment.

In our case, the crime scene only includes a connected device. However, it can contain interconnected objects that provide crucial information for the investigation. For example, the Dot Sleep Sensor can be connected to the Homni Smart Standby solution containing environmental sensors (temperature, humidity, sound level and brightness) and the Reston solution for calculating the heart and respiratory rate of a person. It will therefore always be relevant to start from the study of the data present in the smartphone, because of its role as a catalyst for information. Depending on the purpose of the survey and the need for specific information, objects may be subject to further study. In fact, the probability of finding relevant and complete information in the sensor is lower than in the HCI.

6 Collection and preservation challenge of local objects

The collection of digital evidence comes under the jurisdiction of three staple principles: the relevance, the sufficiency of the data acquired for the survey, and its reliability, especially in terms of verifiability and repeatability of the processing. The evidence must be admissible in court. The ISO/CEI 27037:2017 establishes guidelines for the collection, acquisition and preservation of digital evidence.

Beforehand, the local devices must be isolated from the network when no relevant data is lost due to this action and no malfunction occurs in the system. The investigator dissociates the local network from the crime scene of the outside network by removing the incoming and outgoing connections. This operation requires the physical disabling of connections to the access point or external network port. This process may be supplemented by frequency jamming after a prior impact study of the system. Several difficulties are perceptible in the context of an IoT infrastructure: the question of the network topology and the management of volatile memories. The influence of the reconfiguration of the local network during the collection must also be studied. The investigator can rely on the information collected during the identification phase, particularly on the protocols, the role of the different connected objects and the known dependencies. The local IoT infrastructure is structured according to mesh, star, cellular and broadcast network topologies. Moreover, the collection also depends on the criminal circumstances and the particularities of the crime scene.

Most connected objects studied are not encrypted. The acquisition of data is possible by reading the memory directly. However, during the collection of connected objects on the crime scene, it is possible to perform several acquisitions of data stored via a software approach, especially in the operation of services offering remote access (Secure Shell protocol, Telnet, File Transfer Protocol, server Web, etc.). Deciding a method can be determined by scanning the object. For example, a data dump for the Orvibo kit may be possible via Telnet access or for Philips Hue via API. For some versions of Amazon Echo, Android Debug Bridge debugging can be enabled. Thus, data mining relies on traditional methods and tools for the digital forensics.

7 Conclusion

The rapid growth of IoT inserts new safety and forensic challenges. In front of massive amount of heterogeneous plausible evidence, the digital investigators have to develop digital forensics procedures in order to seriously consider this new field of investigation. Indeed, he must give priority to their approach and operational status with the intention of resolving the riddle of crime quickly.

This article discusses the difficulties of identifying locally connected objects and prioritizing that selection of evidence within the IoT infrastructure. Collection topic will be the focus of a future study. That being the case for digital investigation, the IoT environment is really a source of potential evidence; it's a brand new challenge throughout the forensic science.

References

[Aernouts et al. 2018] Aernouts, M., Berkvens, R., Van Vlaenderen, K., Weyn, M. : "Sigfox and LoRaWAN Datasets for Fingerprint Localization in Large Urban and Rural Areas"; Data, 3,2, 2018, 13.

[Altini et al. 2010] Altini, M., Brunelli, D., Farella, E., Benini, L. : "Bluetooth indoor localization with multiple neural networks"; Wireless Pervasive Computing (ISWPC), 5^{th} IEEE International Symposium, 2010, 295-300.

[Blumenthal et al. 2007] Blumenthal, J., Grossmann, R., Golatowski, F., Timmermann, D. : "Weighted centroid localization in zigbee-based sensor networks."; Intelligent Signal Processing, WISP International Symposium on. IEEE, 2007, 1-6.

[Chan et al. 2006] : Chan, L. W., Chiang, J. R., Chen, Y. C., Ke, C. N., Hsu, J., Chu, H. H. : "Collaborative localization: Enhancing wifi-based position estimation with neighborhood links in clusters"; International Conference on Pervasive Computing. Springer, Berlin, Heidelberg, 2006, 50-66.

[Copos et al. 2016] Copos, B., Levitt, K., Bishop, M., Rowe, J. : "Is Anybody Home? Inferring Activity From Smart Home Network Traffic"; Security and Privacy Workshops (SPW), IEEE, 2016, 245-251.

[Corral et al. 2008] Corral, P., Peña, E., Garcia, R., Almenar, V., Lima, A. D. C. : "Distance estimation system based on ZigBee"; CSEWORKSHOPS'08. 11^{th} IEEE International Conference, 2008, 405-411.

[Dorsemaine et al. 2015] Dorsemaine, B., Gaulier, J.-P., Wary, J.-P. : "Internet of things: A Definition & taxonomy"; Next Generation Mobile Applications, Services and Technologies, 9^{th} IEEE International Conference, 2015, 72-77.

[Forno et al. 2005] : Forno, F., Malnati, G., Portelli, G. : "Design and implementation of a Bluetooth ad hoc network for indoor positioning"; IEE proceedings-Software, 152,5, 2005, 223-228.

[Hegarty et al. 2014] Hegarty, R., Lamb, D. J., Attwood, A. : "Digital Evidence Challenges in the Internet of Things"; INC, 2014, 163-172.

[Höcker et al. 2010] Höcker, M., Berkhahn, V., Kneidl, A., Borrmann, A., Klein, W. : "Graph-based approaches for simulating pedestrian dynamics in building models"; eWork and eBusiness in Architecture, Engineering and Construction, 2010, 389-394.

[Kent et al. 2006] Kent, K., Chevalier, S., Grance, T., Dang, H. : "Guide to integrating forensic techniques into incident response"; NIST Special Publication, 10,14, 2006, 800-86.

[Kirk 1974] Kirk, P. L. : "Crime investigation"; Wiley, 1974.

[Kumar 2014] Kumar, V. : "Effect of environmental parameters on GSM and GPS"; Indian Journal of Science and technology, 7,8, 2014, 1183-1188.

[Lassabe et al. 2009] Lassabe, F., Canalda, P., Chatonnay, P. : "Geolocalisation WiFi et modeles de prediction de la mobilite dans les reseaux multimedia"; 2009.

[Locard 1920] Locard, E. : "L'enquête criminelle et les méthodes scientifiques"; E. Flammarion, 1920.

[Ni et al. 2004] : Ni, L. M., Liu, Y., Lau, Y. C., Patil, A. P. : "LANDMARC: indoor location sensing using active RFID"; Pervasive Computing and Communications, Proceedings of the First IEEE International Conference, 2003, 407-415.

[Oriwoh et al. 2013] Oriwoh, E., Jazani, D., Epiphaniou, G., Sant P. : "Internet of things forensics: Challenges and approaches"; Collaborative Computing: Networking, Applications and Worksharing (Collaboratecom), 9^{th} IEEE International Conference, 2013, 608-615.

[Perumal et al. 2015] Perumal, S., Norwawi, N. M., Raman, V. : "Internet of Things (IoT) digital forensic investigation model: Top-down forensic approach methodology"; Digital Information Processing and Communications (ICDIPC), Fifth IEEE International Conference, 2015, 19-23.

[Rahman et al. 2017] Rahman, K. S., Bishop, M., NSA, A. H. : "Internet of Things mobility forensics"; Digital forensics articles and research papers, 2017.

[Robyns et al. 2017] Robyns, P., Marin, E., Lamotte, W., Quax, P., Singelée, D., Preneel, B. : "Physical-layer fingerprinting of LoRa devices using supervised and zero-shot learning"; Proceedings of the 10^{th} ACM Conference on Security and Privacy in Wireless and Mobile Networks, 2017, 58-63.

[Romero et al. 2009] Romero, H. P., Remley, K. A., Williams, D. F., Wang, C. M. : "Electromagnetic measurements for counterfeit detection of radio frequency identification cards"; IEEE Transactions on Microwave Theory and Techniques, 57,5, 2009, 1383-1387.

[Salman and Jain 2015] Salman, T., Jain, R. : "Networking protocols and standards for internet of things"; Internet of Things and Data Analytics Handbook, 2015, 215-238.

[Sen.se 2018] Sen.se : "Mother and the Motion Cookies", 2018.

[Tedjini et al. 2012] Tedjini, S., Perret, E., Vena, A., Kaddour, D. : "Mastering the electromagnetic signature of chipless RFID tags"; Chipless and conventional radio frequency identification: Systems for ubiquitous tagging. IGI Global, 2012, 146-174.

[Zawoad et al 2015] : Zawoad, S., Hasan, R. : "Faiot: Towards building a forensics aware eco system for the internet of things"; IEEE International Conference on Services Computing (SCC), 2015, 279-284.