# Development of an Advanced Privacy-aware IoT Forensics Process Model

A Dissertation
Presented in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
with a
Major in Computer Science
in the
College of Graduate Studies
University of Idaho
by
Nawaf Abdualaziz Almolhis

Major Professor: Michael Haney, Ph.D.
Committee Members: Terence Soule, Ph.D.; Constantinos Kolias, Ph.D.;
Abdullah Sheneamer, Ph.D.
Department Administrator: Terence Soule, Ph.D.

May 2021

## Authorization to Submit Dissertation

This dissertation of Nawaf Abdualaziz Almolhis, submitted for the degree of Doctor of Philosophy with a Major in Computer Science and titled " **An Advanced User Privacy Aware IoT Forensics Model**," has been reviewed in final form. Permission, as indicated by the signatures and dates below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor: ____________________ Date: 03/10/2021
Michael Haney, Ph.D.

Committee Members: ____________________ Date: Mar. 12 2021
Terence Soule, Ph.D.

____________________ Date: 03/12/2021
Constantinos Kolias, Ph.D.

____________________ Date: 03/12/2021
Abdullah Sheneamer, Ph.D.

Department Administrator: ____________________ Date: Mar. 12 2021
Terence Soule, Ph.D.

## Abstract

Internet of Things (IoT) technologies sense people private information posing a new level of threat to individuals' privacy. Conventionally, consumers have to take some actions to put their privacy at stake, but IoT devices are collecting people's private data without them even noticing. Similarly, when investigating compromised IoT technologies, practitioners have to recover those private information from IoT devices, for evidence; mostly also without the consent of consumers. Digital forensics tools are capable of retrieving data the user considers not present in devices. This leads to a controversial debate on the need for strong privacy measures in the context of IoT forensics. This research aims at finding solutions to this highly fundamental issue that exposes user sensitive information that may go up to threatening lives of the IoT users. In this research, to explicate the problem more technically, a running example that maps the ISO 20137 standard to the IoT ecosystem is presented in Chapter 2. In Chapter 3, a set of requirements that have to be fulfilled by IoT forensics models to preserve user privacy is presented. In Chapter 4, an IoT forensics model is developed by combining features in existing models proposed for different digital forensics domains. Finally, in chapter 5, a searchable encryption scheme that preserves user privacy throughout the process of IoT forensics is proposed.

## Acknowledgements

First of all, I want to thank my GOD (Allah), the Almighty, for the completion of my dissertation. Also, I would like to thank many individual that who support and helped to make this work accomplished its goal.

I would like to express my deepest gratitude to my Major Professor Prof. Michel Haney for his exceptional support, gaudiness, caring, patience, and providing me with an excellent atmosphere for doing this research. I am very grateful to my committee: Prof. Terence Soule, Prof. Constantinos Kolis, and Prof. Abdullah Sheneamer, for their engorgement and insightful comments.

Furthermore, I am thankful for Saudi Cultural Mission and Jazan University for funding my research. Finally, I am deeply grateful to university of Idaho's computer science faculty, staff, and students for how they have affected my life and studies and providing me with a good atmosphere that enabled me to study.

## Dedication

I would like to dedicate this dissertation to praising God (Allah) and asking for his blessing and peace on all of those who contributed to complete this research and making it useful knowledge. Also, this research is detected to:

My Father: Abdualaziz Almolhis,

My Mother: Ameera Obaid,

My Wife: Areej Aloloy,

My kids: Abdualaziz and Btaal,

My Brothers: Dr. Sultan, Abdullah, Ahmed, and Omar,

My Sisters: Gofran, Nosiba, Abeer, Gayda, Kadeja, and Aisha,

Who have been a constant source of support, inspiration, and engorgement over what I have been through during my studding and my life. I am extremely grateful for having you all in my life.

## Table of Contents

## List of Tables

## List of Figures

# Chapter 1: Introduction

## 1.1 Overview

In many countries, privacy is considered a basic human right [1, 2]. Gartner, a non- profit research organization, believes that there will be 20 billion Internet of Things (IoT) devices by 2020 [3]. IoT integrates various objects that are capable of communicating with each other without human intervention. As such, an IoT device accumulates its consumer's personal data and sends it indirectly into cyberspace. This violates the control that consumers should have on their data. When these IoT devices collect and share personal information, consumers become very concerned about the security of the devices and privacy of their personal information [4]. Compromising IoT devices has, however, advanced increased violations of information privacy. Similarly, a situation analogue to consumer privacy, breached by compromising IoT devices, exists in digital forensics investigations involving in IoT devices [5].

Cybercrimes with the power of IoT technology can cross virtual space to threaten human life, along with the increasing number of these crimes, are considered two of the main reasons why we need IoT forensics [6, 7]. However, adopting existing digital forensics solutions in investigations pertaining to IoT devices has raised concerns due to a number of IoT characteristics [8]. Subsequently, in the literature of IoT forensics there are number of challenges that have been encountered by investigators. Among these challenges, the issue of privacy sits at the peak and is an almost untapped research area[9]. That is because, when collecting forensic evidence to reconstruct and locate an incident that involves IoT devices, consumers' privacy may be severely violated [10]. For instance, once the private data is exposed, it is impossible to 'undo' the effects of its exposure should the suspect be found innocent [11]. Similarly, using an IoT device that is connected to a cloud, it would be very difficult for an investigator to collect evidence about a particular perpetrator without breaching the privacy of other cotenants who are using that same cloud [12].

Considerable efforts have been made towards the development of digital forensics solutions in the IoT paradigm [13-15]; yet most of the current solutions have neglected the need for ensuring individual consumer privacy throughout the investigation phase [10, 16]. Some efforts have been made towards the development of privacy-aware IoT forensics solutions [17-19].

However, these solutions have serious limitations. For example, according to [16], the integration of privacy with a conventional digital forensics model, in order to protect consumer privacy in investigations pertaining to such a dynamic IoT environment, is one of the limitations of the said solution. It should be noted that some researchers suggest that this solution does not mitigate the IoT forensic privacy issue and, hence, is not suitable for IoT devices [20, 21]. Similarly, this solution does not give pre-incident preparation measures that take into consideration protection of consumer's privacy before employment of IoT devices. The solution also does not provide live forensics processes (such as the isolation of user data, the location of user data in a cloud, live acquisition, etc.) that are needed to protect privacy while collecting volatile data from IoT devices [8]. In this light, this research will, identify requirements needed to be accomplished by an IoT forensics model that protects privacy. Subsequently, based on those requirements, a model that takes privacy into consideration throughout the process of IoT forensics is proposed.

## 1.2 Problem Statement

Due to the infancy of the adoption of Internet of Things technologies, there has been a lagging development of dedicated digital forensics tools and technologies that can be used in investigations pertaining to compromised IoT technologies. Conventional digital forensics processes and procedures have failed to be employed in IoT forensics. Hence, standard tools and techniques that can be used by investigators ought to be sought quickly. As a result, a number of researchers have proposed solutions in order to reduce this gap. However, these solutions have mostly not taken into consideration aspects of consumer privacy that can be perpetrated by the time of the collection of evidence from IoT devices. An IoT forensics model that protects consumer privacy throughout the phases of investigation is yet to be developed. It is therefore strongly believed that there should be a model, which can be used by practitioners to protect consumer privacy throughout the course of digital investigations involved in IoT technologies, to contribute to the standardization of IoT forensics. The next section presents the main research question that needs to be answered by this research.

## 1.3 Research Question

As the overall goal of this research is to develop an IoT forensics model with consumer privacy in mind, the research question that have to be answered by this research is formulated as follows:

- *How can a consumer privacy-protecting IoT forensics model be developed by integrating existing models, including incorporating views of forensics practitioners?*

## 1.4 Scope of the Research

The research has its scope as follows:

- The research has its focus on the forensics of both cloud connected and standalone IoT devices.

- The research adds measures that protect user data privacy in the investigations IoT ecosystems

- The research uses business process and notation language to represent the proposed model

- The research does not consider the military perspective, rather takes into consideration the business and law enforcement perspectives.

- The proposed model is designed based on according to the Saudi Arabian practitioners' best practices.

### 1.5 Significance of the Research

This research will focus on the development of a model that will protect the privacy of consumers when their IoT devices are involved in a digital forensics investigation. The consumer can be an individual consumer or an organization that has deployed smart IoT devices on its premises. The anticipated significance of this research is that it will come up with a holistic IoT forensics model that takes into consideration the importance of consumer data privacy a way before the adoption of IoT devices by equipping measures that have to be taken from the consumer as well as the provider side of the IoT paradigm. Subsequently, this holistic model proposes live and post-mortem procedures that an investigator should follow in order to conduct a privacy-aware investigation that is a part of the course of investigations involving IoT technologies. In this research, a step-by-step approach that demonstrates the employment of the proposed IoT forensics model will be discussed. These steps will include a means of assessing different IoT technologies with regard to their readiness to preserve the privacy of consumer data in cases where they are involved in a digital investigation [22, 23].

### 1.6 Research Methodology

Research involving the creation of reliable, true, and useful artifacts uses Design Science Research (DSR) as an approach that can guide researchers in developing and evaluating new artifacts [24, 25]. A number of DSR models exist that have been developed in order to get a suitable approach that can be used by computer science researchers. However, some are high-level models that are difficult to follow [26, 27], and some have only focused on the design and development of the artifact, while leaving behind the importance of theory building [26].

Hence, the well-known and highly accepted model, called the Design Science Research model (DSRM) by [28], will be adopted. This model has been used in previous researchin the development of some digital forensic process models, for the reason that it provides easier-to-follow low-level steps compared to other DSR models. In each of the phases of the DSRM model, the mixed research methods approach will be applied. As highlighted in Figure 1.1, all five phases of the model will be accomplished in order to achieve the aforementioned research objectives.

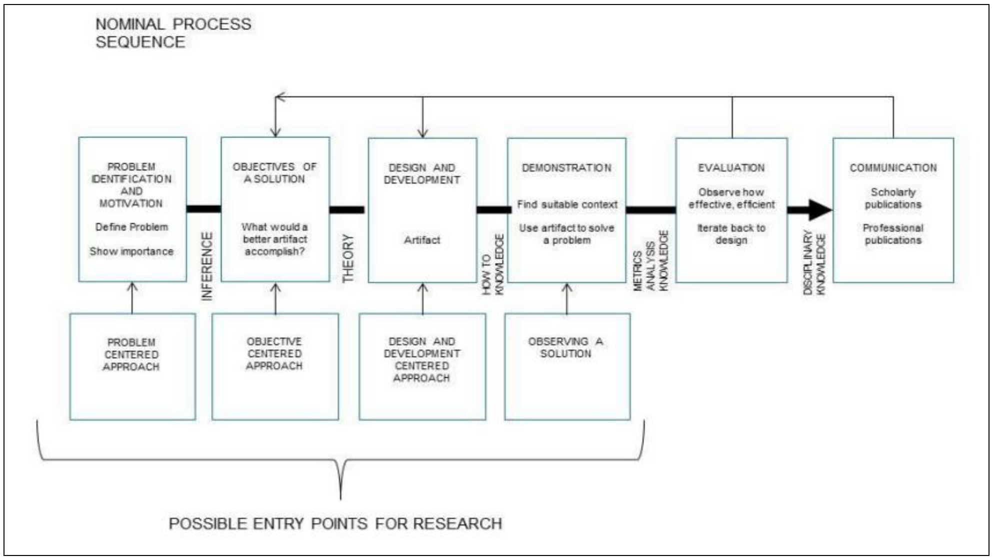
NOMINAL PROCESS SEQUENCE
PROBLEM IDENTIFICATION AND MOTIVATION
Define Problem
Show importance
INFERENCE
OBJECTIVES OF A SOLUTION
What would a better artifact accomplish?
THEORY
DESIGN AND DEVELOPMENT
Artifact
HOW TO KNOWLEDGE
DEMONSTRATION
Find suitable context
Use artifact to solve a problem
METRICS ANALYSIS KNOWLEDGE
EVALUATION
Observe how effective, efficient
Iterate back to design
DISCIPLINARY KNOWLEDGE
COMMUNICATION
Scholarly publications
Professional publications
PROBLEM CENTERED APPROACH
OBJECTIVE CENTERED APPROACH
DESIGN AND DEVELOPMENT CENTERED APPROACH
OBSERVING A SOLUTION
POSSIBLE ENTRY POINTS FOR RESEARCH

Figure 1. 1 Design Science Research Method (DSRM) [28]

*Problem identification and motivation*: The adoption of IoT technology is considered a groundbreaking change that will significantly affect our conventional environment, changing it into a smart environment. A set of planning and strategy, from either the technological, financial or relational perspective, is therefore a prerequisite to anticipate the challenges that come with the use IoT devices. This becomes a strong motivation for this research to investigate not only the capability of an IoT environment for preserving forensics evidence, but also to explore the extent of the consumer privacy breached when conducting IoT forensics.

*Define the objectives for a solution*: Research objectives supposed to be achieved in this research are formulated in this step.

*Design and development*: Concepts and views from digital forensics practitioners and the various literature, such as digital forensic models, the Live forensic model, the Cloud forensic model, and existing IoT forensics models, will be reviewed and combined into a comprehensive workflow in order to lead to a smooth privacy-aware IoT investigation. This workflow prescribes a step-wise approach, from planning of the adoption of IoT devices to its implementation in investigating IoT devices, and within which there are also specific procedures to manage consumer privacy protection that comes with different types of IoT technologies.

*Demonstration:* This model will then be applied to plausible cases that are either formulated by some Saudi Arabia digital forensics experts or practitioners; this will be done in order to study its applicability to security incidents occurring inside the smart environment in Saudi Arabia. To gather a deeper understanding of the hurdles of ongoing processes and the drivers to implement the IoT technologies, semi-structured interviews with the forensics experts and IT divisions will be conducted.

*Evaluation*: The efficacy of this model will be evaluated along with its implementation within these cases as demonstrated in the preceding step. That is, the approaches and processes suggested within the model and how they have been implemented in the cases will be presented to the forensic practitioners who have contributed to the model development (in Step 3). Subsequently, the practitioners will give their feedback on the model. With those insights the model will be modified accordingly so as to achieve its final version.

*Communication*: The design and realization of this model will be documented within the thesis report. Some information may be restricted and therefore will not available to the public. This step also includes publication of the research so as to include it into the domain of knowledge.

# Chapter 2: Mapping Forensic Standard ISO/IEC 27037 to IoT

“Almolhis, Nawaf, and Micheal Haney. "IoT Forensics Pitfalls for Privacy and a Model for Providing Safeguards." 2019 International Conference on Computational Science and Computational Intelligence (CSCI). IEEE, 2019.”

## 2.1 Introduction

In digital forensics, there are conflicting policies; one example is that the data owner has no legal ground to prevent digital forensic investigators from accessing his private data. Similarly, under any privacy law, the data owner has a right to know about or ask for the protection of his private data [29]. However, although forensics investigators may in many instances gather information that is not related to a particular crime, they are still expected to collect any data deemed to be relevant to the crime [30]. To create a consensus among these conflicting policies, digital forensics models have to accommodate privacy-preserving forensic investigations. Many research studies have proposed models for this [2, 31-33]. However, these models do not accommodate the breach of privacy that may occur when conducting an investigation into the IoT ecosystem. That is to say, conventional digital forensic models fall short of preserving user privacy in IoT forensics.

The IoT forensics and conventional digital forensics share a common foundation of privacy perspectives; there are many conventional digital forensics privacy-preserving methods that may be adapted to IoT forensics. However, since IoT forensics has specific privacy-related issues and, therefore, unique privacy-preserving practices must be required for IoT forensics. This chapter discusses privacy issues that come with investigating the IoT by mapping the digital forensics model proposed in ISO/IEC 27037 [34] to the IoT. The ISO/IEC 27037 best practice was proposed in 2012 and was approved by experts in 2018. This chapter is organized as follows: Section 2 presents the related work; in Section 3 the IoT forensics is defined; Section 4 presents the challenges and issues pertaining to IoT forensics: Section 5 discusses the ISO 27037 guideline relative to the IoT architecture; and, finally, Section 6 concludes the chapter.

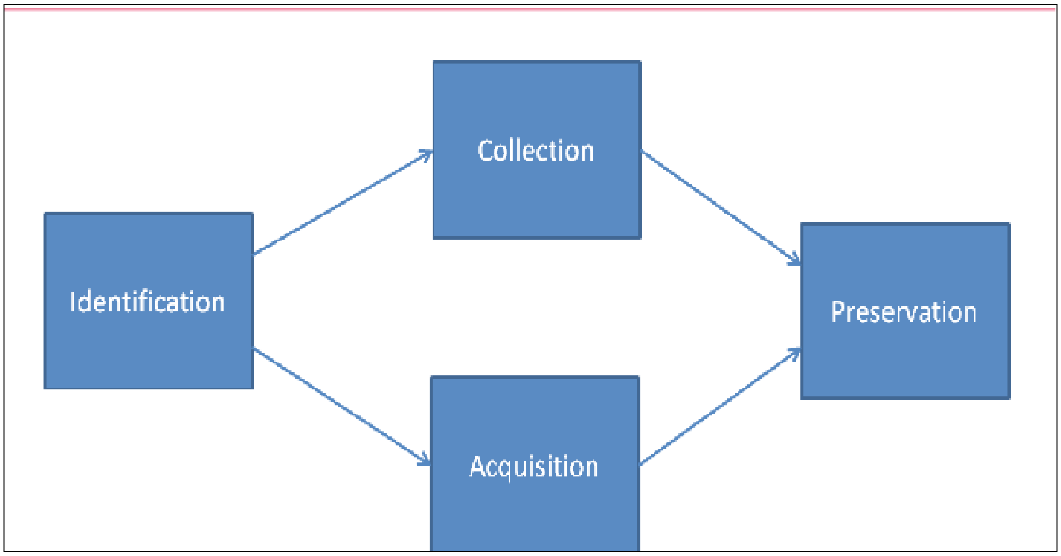

Figure 2. 1 ISO 27037 Model

## 2.2 Related Work

The IoT is a novel internet-based technology that integrates several collaborative and communication technologies to collect personal data. IoT technologies have taken root in all sorts of lives of individuals and society at large. For example, IoT architectures include tracking and identification technologies that are integrated through actuator and sensor networks to allow communications between distributed intelligent objects, and include wearable smart devices that may collect information about locations without the consent of the user [35]. However, the collection, handling, and distribution of such personal data present unique ethical challenges in terms of user privacy. Ethically, these kinds of activities may expose information users would have kept to themselves unless they were forced to reveal this information to others [35]. As a result, from the perspective of privacy, the consumer data protection of IoT technology can be questioned because the collection and use of IoT generated data may occur without the consent or permission of its users. Information privacy can be referred to as "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" [36].

There are three information privacy concerns that can be attributed to consumers. These include the very act of collecting consumer data, the control consumers may have over collected data, and the consumers' awareness of how collected data are used [36-38]. In this chapter, the term "data" is used synonymously with the terms "information" and "evidence." The collection of personal information has to be done only on the basis of an agreed-upon social contract. Similarly, collected personal information can be used by a third party when the consumer is granted control to have a say about it, i.e., approval, modification, or opting for an exit. Finally, consumers should be aware of and be informed about the consequences of giving away the stewardship of private information. Making consumers aware of the collection of private data should add transparency and fairness to the act of collection.

Many efforts have been made by researchers to overcome the IoT privacy issues. Privacy-enhancing techniques has been employed for preserving privacy in the IoT ecosystem by simply removing user identifying information from data (anonymization techniques). The state of the data set is changed in a way so that no consumer is identified as being the owner of the data. Another example is utilizing security techniques for privacy-preserving mechanisms in the IoT ecosystem. Well-known security techniques include the employment of an encryption technology to ensure confidentiality, integrity, and the availability of data collected by IoT technologies [39, 40].

There are a number of privacy-preserving or enhancing technologies in the literature. However, in this chapter the authors will try to present the most-cited privacy-preserving technologies proposed for the IoT ecosystem. Substitution is an anonymity technique that randomly replaces identifying or quasi-identifying information from the data; for example, substituting the string "XA5Y" for the name "Bob" or replacing the string "Peter," which looks like legitimate data, for the name "Bob" [41]. According to [41, 42], there are some other technologies with similar patterns to substitution, such as masking and nulling out.

Shuffling arranges values in columns in a way that displaces associations between attributes in order to produce an anonymous dataset for the end-user. Shuffling works at the application level with the goal of removing sensitive attributes and identifying attributes without altering the entries in any other way [39, 42]. Shuffling is primarily an anonymization technique that works best in preserving privacy in data duplication [43]. IoT forensics that collects or duplicates evidence from IoT ecosystems could utilize this technology. Sampling is a technique that releases only partial data from a complete data set by suppressing some of the data; the released data appears to represent the whole data set. Sampling is an anonymization technique and is applied at the application layer for categorical and numerical data [39].

Differential privacy cleverly manipulates sequence queries so as to protect aggregated statistical data needed by attackers to deduce an individual's private data. Differential privacy intentionally introduces noise into the sequence of queries in order to obfuscate the presence or absence of individual data in a data set [44]. This technique can be applied to evidence collected from an IoT ecosystem before the analysis and examination activities start to obscure association between evidence and innocent users of the IoT technology [45]. The generalization technique is applied at the sensor and application level of the IoT ecosystem to decrease the granularity of some attributes in the data [42]. Generalization is an anonymization technique that best works with the protection of categorical and numerical information.

Encryption completely hides the information with the use of a cryptographic key to achieve anonymity and is widely used at any stage of data collection. Encryption has its own challenges in digital forensics, and hence, its usage for privacy goals may add to the already existing issues of IoT forensics by jeopardizing the process of investigation because the forensics investigators would likely miss important evidence of the case [45]. Blockchain is a well-distributed cryptographic system and its applications in IoT forensics for privacy preservation have been investigated by a number of researchers. Blockchain offers anonymization for numerical data or categorical information, such as the identity of the consumer [20, 46, 47].

Mix networks provides for the anonymity of senders and receivers within a network by using anonymizing encryption techniques that can help protect categorical and numerical information in the network layer [39]. Onion routing also provides sender and receiver anonymity through socket connections by using indirect paths that add layer upon layer of encryption for each hop. Onion routing has been employed in a number of digital forensic models for privacy preservation [45]. Privacy-enhancing techniques discussed in this section are reviewed thoroughly to help find an applicable solution for the different processes of IoT forensics for privacy requirements. The next section defines IoT forensics in relation to digital forensics.

## 2.3 IoT Forensics

Digital forensics investigations usually hinge on the admissibility of digital evidence. According to Grobler (2011), digital evidence is defined as “information of probative value stored or transmitted in a digital form”. Researchers use a number of terms, including “Digital Forensics” or “Data Forensics,” when referring to the digital forensics’ science. However, the use of “Digital Forensics” is very familiar among researchers, forensics experts, and digital forensic first responders. Digital forensics is defined as “the practice of scientifically derived and proven processes and tools applied towards the recovery, preservation and examination of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of the events as forensic evidence helping to anticipate the unauthorized actions shown to be disruptive to planned operations" [48-50]. According to this definition, in digital forensics, both tools to be utilized and processes to be followed when conducting an investigation have to be approved by a court of law.

Similar to other domains, like Computer Forensics, Cloud Forensics, and Network Forensics, IoT Forensics is not different and can be considered as part of digital forensics science. Each domain is expected to have specialized tools and models for investigating security incidents pertaining to respective technologies [51]. IoT forensics is a cross-discipline between IoT and digital forensics and can be referred to as the application of digital forensics principles and procedures in an IoT ecosystem. In [20], IoT forensics is also defined as the process of collecting, analyzing, storing, and presenting digital evidence within IoT devices in a legally binding manner.

In the literature, the phrase "digital forensic investigation" is very common, and researchers use it for different situations. For instance, Henry et al. [52] from the SANS Institute claim that digital forensic investigation refers to a process of investigating cases such as employee abuse of resources, espionage, or financially motivated attacks. In [53], digital forensics investigation is referred to as "a process that develops and tests hypotheses to answer questions about an event that occurred."

Similarly, in [54], researchers consider digital forensics investigation as a process that examines and analyzes collected digital evidence for the establishment of factual information for judicial review. Kohn et al. [55] also define digital forensics investigation as a special type of investigation that uses scientific procedures and techniques for digital evidence to come up with results that are admissible in a court of law.

In the two latter instances, digital forensics investigation is related to cases where their results will be submitted to a court of law, as legal aspects of digital forensics, while the two former instances do not explicitly mention that digital forensics investigation answers questions of criminal activities, but rather for due diligence purposes. However, the four definitions together indicate that digital forensics investigation can be used in an incident caused by events that violate organizational policies or events that violate local or international laws. Therefore, this chapter defines the phrase IoT forensics investigation as a process that develops and tests a hypothesis to determine and relate digital evidence extracted from an IoT ecosystem to establish factual information about security incidents caused by violating policies and/or laws.

Currently, because of the infancy of the field, there is no single IoT forensics standard model or tool in the literature [46, 51]. However, there are some substitute approaches that can be used to qualify IoT forensics process models. For example, digital evidence collected from an IoT environment can be related to the well-known four important principles most used and published in the “Good Practice Guide for Computer-Based Electronic Evidence” [56]. As a result, the following four principles can be used for the admissibility of evidence collected from IoT devices:

- No action taken by law enforcement agencies or their agents should change data held on an IoT device that may subsequently be relied upon in court.
- In exceptional circumstances, when a person finds it necessary to access original data held on an IoT device, that person must be competent to do so and be able to provide evidence explaining the relevance and the implications of his actions.
- An audit trail or other records of all processes applied to digital evidence residing in an IoT ecosystem should be created and preserved. An independent third party should be able to examine those processes and achieve the same result.
- The person in charge of the investigation (the case officer) has overall responsibility for ensuring that the law and these principles are adhered to.

Consequently, IoT forensics investigators and IoT forensics tools developers must closely follow these four de-facto principles. Subsequently, the terms and rules given in the four principles have to be upheld.

### 2.4 Issues and Challenges in IoT Forensics

The IoT is growing rapidly and being deployed in a wide range of applications, from smart grids to healthcare and intelligent transport systems. The sensitive nature of private data that are collected and transmitted by IoT devices have attracted nefarious perpetrators [57]. IoT devices are highly constrained in terms of memory and processing for being secured properly; their inability to support conventional security measures make them a low-hanging fruit for exploitation [8]. For instance, the Mirai attack in 2016, which is a prime example of a such a notorious attack against the IoT ecosystem, was a malware that infected IoT devices and further used them as a launching platform for DDoS attacks [51, 58].

IoT security attacks may target and exploit vulnerabilities in any of the components of the IoT ecosystem: the sensor, the edge/fog, or the cloud layer. Digital forensics is needed to hold perpetrators accountable [59]. However, digital forensic practitioners face challenges in investigating IoT devices [14]. These challenges may include the following.

Scaled-up heterogeneous devices, such as different types of IoT technologies, operating systems, and network protocols from different vendors, make it impossible to have a standardized approach in device-level forensics. For instance, device heterogeneity in the IoT ecosystem may call for specialized evidence collection tools [10]. For communication purposes, different wireless protocols may be associated with different IoT devices. Collecting evidence from ZigBee-enabled devices, which need a smart hub to connect to the Internet, and collecting evidence from WiFi-enabled devices that may directly connect to a gateway router may not be the same. As a result, this makes it difficult to identify the source of the evidence in the IoT ecosystem.

*Data management in devices:* the heterogeneous nature of devices regarding the methods in which data is distributed, aggregated, and processed pose challenges in forensic investigations. Such highly complex data collected from different sources in the IoT may hinder the performance of the smooth analysis needed to make the required decisions [16].

*User privacy:* collecting evidence only from suspected devices in a way that preserves the privacy of other innocent users poses challenges to investigators. Not only during evidence collection, but privacy is an important factor to consider when analyzing and correlating the collected evidence that may contain personal information. Privacy-preserving measures have to be taken throughout the course of IoT forensics procedures [10, 16, 60]. In the literature, it is obvious that IoT consumers lack the knowledge and understanding of their rights regarding their personal data. Most of the researches and model development for the preservation of privacy of consumer data are focused more on post-mortem or passive preservation. Hence, the development of a holistic model is necessary, one that provides consumers a means of protecting their private data *before* the deployment an IoT system.

*Volatile data:* typically this is produced through some type of wireless network, such as the WiFi, Bluetooth, or ZigBee. The volatile network traffic sent and the corresponding data from the cloud by the IoT devices pose challenges to investigators.

*The volume of data:* the volume of data involved in IoT technologies creates another challenge to IoT forensics, especially in situations where considerable time has elapsed after the incident has occurred. For example, the volume of IoT data captured by sensors and smart devices from networks and the cloud complicates the identification of relevant data.

*Dependency on a cloud service provider:* collection of evidence residing at the cloud data center frequently relies on the service provider. This could create mistrust of the provider, who might manipulate evidential data in order to preserve the reputation of his cloud services [8].

## 2.5 Mapping ISO/IEC 27037 to the IoT Architecture

ISO 27037 is the first international standard organization to try to institute a common baseline for the process of conducting digital forensics. The guideline provides initial steps for the digital forensics process, including identification, collection, acquisition, and preservation. However, there is a lesser chance of breaching private information if privacy-preserving measures are strictly taken into consideration throughout the course of these four steps. Thus, other remaining steps, such as examination and analysis of IoT forensics, are more or less the same as the conventional practices and hence are not considered in this chapter.

The identification step, according to the guideline, is the search, recognition, and documentation of sources of potential evidence by prioritizing these sources based on the volatility of the evidence and minimizing any damage to any potential evidence. Subsequently, the collection step is a process where devices involved in the suspected crime are removed from their original location to a controlled environment for further acquisition and analysis. The acquisition step is the process of creating an exact copy or image of an item from the potential digital evidence. And finally, the preservation step is about protecting the integrity of the acquired potential digital evidence from tampering and spoilage.

In this mapping process, each of the steps provided in ISO 27037 will be mapped to the three different layers of the IoT architecture of a home system as proposed by [61]. The layers include perception, network, and application layers. The perception layer may include physical and logical sensors and other smart devices used for the real-time monitoring of houses. The network layer is the layer that is responsible for the transmission and processing of the information gathered from the devices at the perception layer; this includes all kinds of network technologies. The network layer is linked to the perception layer by the home gateway. Finally, the application layer consists of the application support platform, which usually resides in the cloud and applications that interconnect the smart devices and users. At each of these layers, different types of digital forensics can be employed. For instance, the application layer involves cloud forensics, the network layer involves network forensics, and the perception layer involves a computer or device forensics. Hence, this mapping will be conducted in this manner.

The guideline proposed in ISO 27037 starts with the identification step. The identification of different sources of digital evidence of the IoT architecture will then be considered. This includes the identification of the evidence sources at the cloud-residing application support platform, the identification of the evidence sources at the network layer, and the identification of the evidence sources the smart devices at the perception layer. The following sections discuss the mapping and focus on privacy issues in each of the steps.

## 2.6 Discussion

### 2.6.1 Identification

When identifying the potential sources of digital evidence, the process starts with the identification of the stand-alone systems involved in the case. Regarding IoT architecture, the stand-alone devices can be found at the perception (smart devices) and network layers (communication devices). Because the application layer, which in other words is the IoT application platform residing in the cloud, it is unlikely feasible for an investigator to access the distributed networked systems of the cloud data center; it is not therefore applicable at this layer.

In this step, the model does not consider any means of preserving individual private information. Hence, in this step, an IoT smart device that continuously collects the private information of the people around it can be mistakenly identified as an important source of digital evidence. Since the identification step is a decision-making process, in which a determination must be made from where to collect and acquire evidence, it may lead to a greater chance that the investigator will collect private information that is not related to the case.

Apart from the smart end-user devices, in a networked environment, the identification of where potential evidence is stored is important. Hence, in the network layer of the IoT architecture, the investigator has to identify possible sources of evidence. For instance, the home gateway that gathers information from the perception layer, and transmits it to the network layer, can be identified as a potential source of evidence. However, sitting at the peripheral, the home gateway may collect information from different smart devices owned by different individuals who are using the same home IoT network. Thus, identifying the home gateway, or any other network device, may lead to the collection of the private data of innocent individuals.

In [62], the authors proposed a six-step IoT evidence source-identification method in which they did not consider the privacy issue that may arise with their process of identification. What are considered as two different steps in the ISO 27037 guideline, identification and collection, is taken as one step and is referred to as IoT device identification in this chapter. The method starts with the identification of the IoT layer in which the device resides. This is followed by establishing access to the device. Next is definition, which defines the data category that the suspected source may provide, and then isolation, which isolates the device from the IoT network. The final step is to ensure the availability of the device for investigators.

The privacy concerns that might be related to the identification of the potential sources of the evidence are mainly dependent on the sensitivity of the investigation being conducted. In cases where the investigation is sensitive and needs to be conducted covertly, investigators must prioritize the potential sources of evidence on the basis of the privacy of the data contained in the case and the relevancy the data may have to the case. Similarly, in cases where the investigation is less sensitive and can be conducted overtly, users of the IoT devices may be interviewed and, on that basis, the investigator may identify sources that may not contain private information of innocent users.

### 2.6.2 Collection

In this step, sources identified in the preceding step as potential sources of evidence are collected and taken to a laboratory. This step of the model also partially involves IoT forensics; the application layer concerns the data-center side of cloud forensics, which is limited only to the service provider. In this step, the investigator might recognize some additional devices that could contain evidence that might help the case. For those devices to be transferred to a digital forensic laboratory, it depends on the state of the device. For example, some of the devices are mission-critical and hence need their data to be acquired live at the scene. In this case, those devices are isolated from the home IoT network to keep new data from contaminating existing data. This occurs at those times when data acquisition runs parallel to the process of collecting items identified as potential sources of evidence. However, if the device does not have mission-critical information and can be or has been switched off, then it must be transferred to the forensics lab. Only after the above can the process of acquisition of evidence commence.

Apart from collecting suspected devices, this step also includes the selection of the forensic tools that is are going to be used for the acquisition of the evidence. Importantly, an investigator must always select tools with privacy-preserving mechanisms. Privacy-preserving mechanisms can in some cases be disregarded when strong measures for the preservation of privacy have been taken in the preceding identification step.

### 2.6.3 Acquisition

One of the most important steps among all steps, when it comes to IoT forensics, is the acquisition of digital evidence. Similarly, this step is very important when conducting a privacy-aware investigation. According to the guideline, the process of acquiring potential evidence can be conducted in either a live or static manner. Live acquisition is the capability of acquiring live evidence from a live system, while static acquisition is a means of acquiring evidential data from a powered-off system. Thus, in most cases in IoT forensics, the live acquisition process is more practical and requires some dedicated methods that can be used particularly for collecting evidence from devices at different layers of the IoT architecture.

The guideline does not propose specific privacy-preserving mechanisms for use in data acquisition. However, the guideline promotes partial acquisition, where only some selected data must be acquired; according to the guideline, this can be performed when the target system storage is too large to be fully acquired, when the system is too critical, or when a search warrant limits the scope of the acquisition. To an extent, the latter two situations support the reasons for protecting private data. But this alone in not enough; a rather more reasoned and practical way of protecting private data in IoT forensics must be considered to circumvent privacy violations.

In conventional digital forensics, for example, researchers like [29] propose a data classification mechanism with which they could define the level of privacy associated with the data to be acquired for investigations. Likewise, in [63], the researchers propose a partitioning and ordering mechanism by which the portion of the data assumed to be more important is checked for its privacy-accuracy level. In addition, in [10], the researchers propose a context-based data collection mechanism; the IoT consumer or the data owner is asked for the existence of private data in order to prevent the requisition of personal data.

### 2.6.4 Preservation

According to the guideline, this step is all about sealing collected devices or acquired data with some verification functions or digital signatures to determine that the digital evidence copies are equivalent to the originals. In this context, the step does not change and would remain applicable for IoT forensics.

## 2.7 Chapter Summary

In this chapter, the privacy issues related to IoT forensics investigations are surveyed. The well-known privacy-enhancing technologies are reviewed, IoT forensics is defined, IoT forensics challenges reported in the literature are discussed, and finally, an approved ISO/IEC 27037 is mapped to the three-layered IoT architecture to show the defects associated with the adoption of digital forensics models to IoT forensics, especially regarding the privacy issues that may arise.

# Chapter 3: Requirements for a Privacy aware IoT Forensics Model

"N. Almolhis, A. Alashjaee, M. Haney, "A review: Requirements for IoT Forensics Models," Transactions on Computational Science and Computational Intelligence, Advances in Security, Networks, and Internet of Things, Springer Nature, 2020." **(Accepted)**

## 3.1 Introduction

The increased use of IoT technologies by both individuals and enterprises for various applications will continue to spread. This has brought about the need for IoT forensics for investigations on security incidents in the IoT ecosystem. Consequently, this growth of usage has resulted in an increase of research in IoT forensics. Several IoT forensics methods and tools have been proposed in the literature and will be reviewed.

There are several issues and challenges investigators encounter when conducting IoT forensics [1]. Many report that conventional digital forensics tools and methods cannot effectively be used in investigating IoT security incidents. In this light, new IoT forensics tools, as well as processes, are urgently required that can guide the procedures of investigators conducting IoT forensics.

In this chapter, the authors review the most prominent technical research publications to date on IoT forensics. Some IoT forensic challenges are identified. Subsequently, because of these challenges, we present a set of requirements that an IoT-forensics process model should address. The IoT-forensics process models identified in the literature have been evaluated based on the deduced requirements to shed light on the gaps that remain.

This chapter is organized as follows: Section 2 presents the related work; Section 3 presents challenges and issues of IoT forensics; Section 4 identifies IoT forensics models requirements; Section 5 maps the requirements to the existing models; and, Section 6 evaluates the requirements. Finally, Section 6 concludes the chapter.

## 3.2 Related Work

The "Internet of Things" or IoT is a novel implementation of internet-based technologies that integrates several collaborative communication technologies, which exist as objects in the real world, as opposed to traditional computers, which exist with a monitor and keyboard for interaction by humans. IoT technologies have taken root in the lives of individuals and society at large, in both small-scale consumer items, such as dishwashers and refrigerators, to larger moving objects, such as automobiles, to even larger-scale industrial systems that make up parts of manufacturing or energy plants. A significant characteristic of these IoT devices is their ability to collect data from many types of sensors. Much of the collected data could be considered personal to the devices' owners. IoT architectures include tracking and identification technologies, integrated through actuator and sensor networks, to allow communications between distributed intelligent objects including, for example, wearable smart devices; these may collect information about locations without the consent of the user [2, 3].

The best practices and guiding principles of digital forensics and any sub-disciplines, such as IoT forensics, require not only the tools used for conducting an investigation but also the process of conducting an investigation that can then be reviewed, calibrated, verified, and approved in such a way that it is independently reproducible. This helps ensure that digital evidence collected in a forensically-sound way will be admissible to a court of law. Hence, to standardize and capture the process of conducting digital investigations, process models are considered vital to for expediting the investigations and addressing issues investigators encounter, especially with new technologies [4, 5]. Widely accepted models point out that different tools are necessary for the investigations of different technologies and need to be developed.

There are some IoT-specific digital forensic models proposed in the literature. For instance, in [6], researchers discussed the issues in digital forensics brought about by the use of IoT systems and proposed an IoT forensics deployment model (M1) that incorporates probable solutions in each of its phases. Researchers focused on the post-mortem investigation; and they did not design and implement the model. Although the issue of user data privacy had been raised, these researchers did not propose any probable solution in this regard.

The model proposed in [7] divides the IoT network into three zones that include internal, middleware, and external networks, as depicted in Figure 3.1. The researchers adopted the triaging principle to their model in conjunction with the zones. They claim that their model (M2) would be suit better suited for internal incident responders. The model does not cover aspects of IoT devices and applications of an IoT-forensics investigation. Rather, it works on the network layer of the IoT ecosystem. Additionally, the model does not take into consideration user privacy issues and, hence, does not provide measures to protect user identity in the live data collected for analysis.

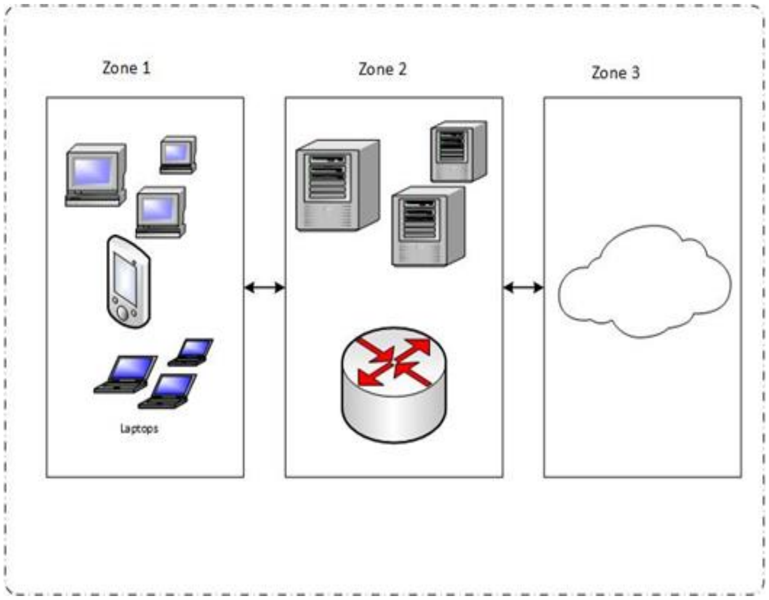

Figure 3. 1 Digital Evidence Acquisition Model for IoT Forensic [7]

The model proposed in [8] is relatively more comprehensive. This model (M3) would increase the level of trust among different IoT consumers. In doing so, the researchers proposed the employment of Blockchain technology with a lightweight fragmented ledger in the IoT infrastructure, as can be seen in Figure 3.2.

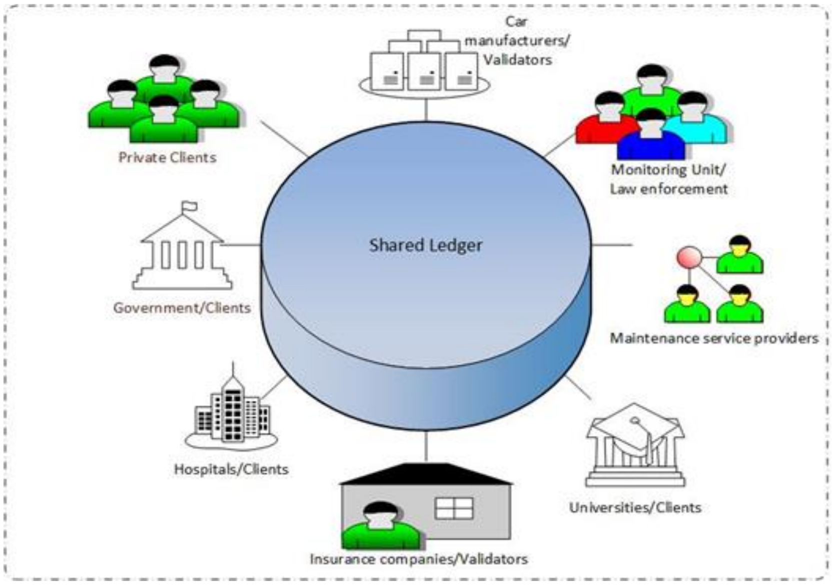
Car manufacturers/ Validators
Private Clients
Monitoring Unit/ Law enforcement
Shared Ledger
Government/Clients
Maintenance service providers
Hospitals/Clients
Universities/Clients
Insurance companies/Validators

Figure 3. 2 Lightweight Blockchain [8]

In [9], researchers proposed a Blockchain-based IoT-forensics model that preserves identity privacy throughout the lifecycle of the evidence collection. They have given a working definition for IoT forensics. Figure 3.3 shows the sequential diagram for the proposed evidence collection. An issue for this model is that it only focuses on the evidence collection after other important phases of the IoT investigation (M4).

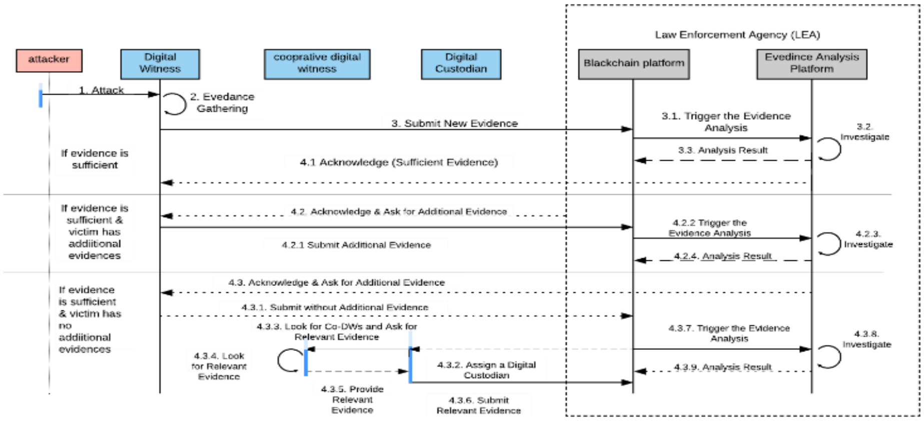
Law Enforcement Agency (LEA)
attacker
Digital Witness
coopratiive digital witness
Digital Custodian
Blackchain platform
Evedince Analysis Platform
1. Attack
2. Evedance Gathering
3. Submit New Evidence
3.1. Trigger the Evidence Analysis
3.2. Investigate
3.3. Analysis Result
If evidence is sufficient
4.1 Acknowledge (Sufficient Evidence)
If evidence is sufficient & victim has addiitional evidences
4.2. Acknowledge & Ask for Additional Evidence
4.2.1 Submit Additional Evidence
4.2.2 Trigger the Evidence Analysis
4.2.3. Investigate
4.2.4. Analysis Result
If evidence is sufficient & victim has no addiitional evidences
4.3. Acknowledge & Ask for Additional Evidence
4.3.1. Submit without Additional Evidence
4.3.3. Look for Co-DWs and Ask for Relevant Evidence
4.3.4. Look for Relevant Evidence
4.3.2. Assign a Digital Custodian
4.3.5. Provide Relevant Evidence
4.3.6. Submit Relevant Evidence
4.3.7. Trigger the Evidence Analysis
4.3.8. Investigate
4.3.9. Analysis Result

Figure 3. 3 Sequential Diagram of Evidence Collection [9]

Researchers defined in [10] a conceptual model (M5) that uses a secure logging scheme that stores evidence in a centralized repository, as shown in Figure 3.4. Most of the work of the researchers focused on the preparation of the proactive part of the investigation. However, in the case of live forensics and most post-mortem processes, what is proposed in the model falls short.

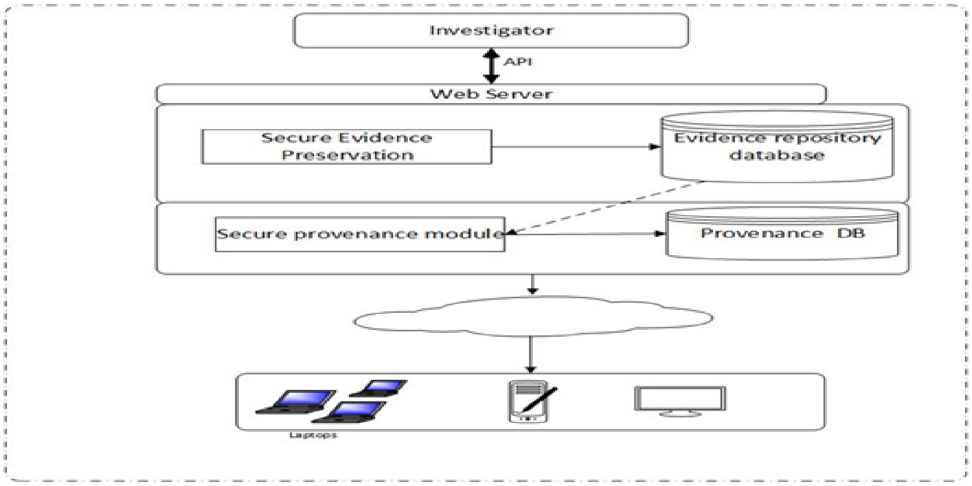

Figure 3. 4 A Conceptual Model for IoT Forensics [10]

In [11], the researchers introduce a fog-based IoT-forensics framework (M6) in which they try to accommodate challenges associated with IoT forensics. The framework uses fog computing to recover forensic evidence from the IoT ecosystem. The framework is proposed as a monitoring tool that identifies malfunctioning devices, collects associated evidence, and analyzes the evidence to determine suspicious activities.

In [12], the researchers proposed a privacy-aware model (M7) that stimulates the cooperation of different consumers in digital forensics. The model promotes a collection of evidence from surrounding IoT devices in order to fully describe the context of a crime scene. However, some researchers have suggested that this solution does not mitigate the IoT-forensic privacy issues and, hence, is not suitable for IoT devices [9]. One of the main issues associated with this model is that it applies an existing conventional digital forensics model directly to accommodate challenges faced by IoT-forensics practitioners.

In [13], the researchers based their model (M8) on three phases, forensic readiness, forensic initialization, and forensic investigation. The researchers did not raise the issue of privacy that most of the IoT-forensic community has pointed to as one of the main challenges that IoT-technology investigators are suffering from. Likewise, the live forensic measures reported in most papers as an integral part of the investigation in the IoT ecosystem are not discussed as a main part of the model. The researchers have raised relatively more general steps, which makes its suitability in IoT forensics difficult. Nevertheless, the three phases proposed in the model are real parts of most investigations.

Most of the solutions in the literature focus on the post-mortem or aftermath investigations rather than a readiness or proactive ones. For instance, if a security incident did not disrupt the whole system, consumers cannot determine the status of a compromised IoT device. It is recommended to have some readiness measures, such as monitoring mechanisms that can premeditatedly collect data, which can be used as forensic evidence [11]. Similarly, extending conventional digital forensic tools and models to the IoT ecosystem is considered inefficient. Hence, a dedicated holistic model is needed that uses the readiness, live, and post-mortem measures together for IoT forensics [14][64]. To that end, it is obvious that researchers are using either research products or their own experience to formulate the models. It is also notable that the models are accommodating several aspects of IoT forensics rather than randomly based on the need or experience of the developer. Thus, to create some sort of uniformity, there must be some essential requirements in the IoT-forensics model that are deemed necessary and are considered suitable in the IoT ecosystem.

## 3.3 Issues and Challenges in IoT Forensics

IoT is growing rapidly and being deployed in a wide range of applications starting from smart grids to healthcare and intelligent transport systems. The sensitive nature of private data that is collected and transmitted by IoT devices have attracted perpetrators [57]. IoT devices are highly constrained in terms of memory and processing to be secured properly, their inability to support conventional security measures make them a low hanging fruit for exploitation [8]. For instance, Mirai attack in 2016, which is a prime example of such a notorious attack against the IoT ecosystem, a malware that infects IoT devices and further uses them as a launch platform for DDoS attacks [51, 58].

IoT security attacks may target and exploit vulnerabilities in any of the components of the IoT ecosystem: i.e., the sensor, the edge/fog, or the cloud layer. Digital forensics is needed to hold perpetrators accountable [59]. However, digital forensic practitioners are facing challenges in investigating IoT devices [14]. These challenges may include:

*Scaled-up Heterogeneous Devices*: Different types of IoT technologies, Operating systems, and network protocols from different vendors make it impossible to have a standardized approach in device-level forensics. For instance, device heterogeneity in the IoT ecosystem may call for specialized evidence collection tools [10]. For communication purposes, different wireless protocols may be associated with different IoT devices. Collecting evidence from ZigBee enabled devices which need a smart hub to connect to the internet, and collecting evidence from WiFi-enabled devices that may directly connect to a gateway router may not be the same. As a result, this makes it difficult to identify the source of the evidence in the IoT ecosystem.

*Data Management in Devices:* Heterogeneous nature of devices regarding the methods in which data is distributed, aggregated and processed pose challenges in forensic investigations. Such highly complex data collected from different sources in IoT may hinder the performing of the smooth analysis needed to make the required decisions [16].

*User privacy:* Collecting evidence only from suspected devices in a way that preserves the privacy of other innocent users is posing challenges to investigators. Not only at the evidence collection but also in Privacy which is an important factor to consider when analyzing and correlating the collected evidence which may contain personal information. Privacy-preserving measures have to be taken throughout IoT Forensics procedures [10, 16, 60, 65]. In the literature, it is obvious that IoT consumers are lacking the knowledge and understanding of their rights in their data. Most of the researches and model developments with privacy preservation of consumer data are focused more on post-mortem or passive preservation. Hence, the development of a holistic model, that provides consumers the means of protecting their private data way before deployment IoT systems, is necessitated.

*Volatile data:* Typically, through some type of wireless network i.e., WiFi, Bluetooth, or ZigBee. The volatile network traffic sent to and the corresponding data from the cloud by the IoT devices poses challenges to investigators.

*The volume of data:* The volume of data involved in IoT technologies is causing another challenge to IoT forensics, especially in situations where considerable time has elapsed after the incident has occurred. For example, the volume of IoT data captured by sensors and smart devices from networks and the cloud complicates the identification of relevant data.

*Dependency on Cloud Service Provider:* Collection of evidence residing at the cloud data center frequently relies on the service provider. This would create unnecessary trust of the provider who may manipulate evidential data for preserving the reputation of his cloud services [2, 8].

Digital forensics models are used as a guide by practitioners as well as digital forensics tools developers. Hence, an IoT forensics model must get solution to the challenges reported in this section. Some of the challenges are directly related to the characteristics fo the IoT ecosystem while some are indirectly related such as those arising from other technologies that IoT device usually use to accomplish daily activities including Cloud Computing, Fog Computing, and Edge Computing.

### 3.4 IoT-Forensics Model Requirements

In an effort to overcome these challenges, the researchers formulated certain IoT-forensics requirements that they think may provide solutions, as depicted in Table 3.1. In [1, 25], researchers presented essential requirement analysis approaches that can help digital forensic readiness (DFR) processes to be successfully implemented in an IoT environment. These include the extraction of digital evidence, parsing forensic logs, digital preservation, creation of hash values, evidence storage, log analysis and characterization, and readiness reporting. Although DFR is considered important in IoT forensics, these researchers did not take into consideration privacy measures that have to be taken at the readiness stage.

The requirements proposed for the digital forensic investigation of the IoT include integrity, non-repudiation, relieve single point-of-trust, persistence of forensic analysis, lightweightness, and privacy [8, 26, 27]. Similarly, some key requirements are proposed in [21] and include managing IoT data volume, mitigation of privacy risks, guidelines for the IoT deployment approaches, and dealing with system identification and human behaviors.

Some privacy-oriented IoT-forensics requirements are proposed in [28]. The researchers translated privacy principles in ISO 29001 into requirements. These include consent and choice for use when an IoT consumer should give consent on the collection of his data. The purpose legitimacy and specification requirement promotes the understanding that the consumer has to be informed about the reason for the data collection. The collection limitation requirement refers to the collection of the data strictly relevant to the case at hand.

The data minimization requirement answers to the large volume of data that can be collected from the IoT ecosystem and promotes the reduction of the data to the minimum volume possible. Based on the use, retention, and disclosure limitation requirement, data collected must not be used for a purpose other than the one originally specified. The accuracy and quality requirements ask that an investigator conduct the data collection process in a trusted and admissible way.

Table 3. 1 Identified Requirements.

| ID | Requirements |
|---|---|
| RQ1 | Digital Forensic Readiness (DFR) processes including extraction of digital evidence, parsing forensic logs, digital preservation, creation of hash values, evidence storage, log analysis and characterization, and readiness report. |
| RQ2 | The volume of IoT data captured by sensors and smart devices from networks and the cloud complicates the identification of relevant data. |
| RQ3 | The parties should be held responsible for their actions by providing proof of integrity. |
| RQ4 | The system should have minimum overhead on endpoints since it includes multiple parties that may have different capabilities and resources. |
| RQ5 | Consent and choice when an IoT consumer should give consent on the collection of his data. |
| RQ6 | The purpose legitimacy and specification requirement promotes recognition that the consumer has to be informed about the reason for the data collection |
| RQ7 | Collection limitation requirement refers to the collection of the data strictly relevant for the case at hand. |
| RQ8 | Data minimization requirement answers to the large volume of data that can be collected from the IoT ecosystem and hence promotes the reduction of the data to the minimum volume possible. |
| RQ9 | The use, retention and disclosure limitation requirement is to assure that data collected must not be used for a purpose other than the one originally specified. |
| RQ10 | The accuracy and quality requirements ask of an investigator that the data collection process must be done in a trusted and admissible way. |
| RQ11 | The openness, transparency and notice requirement, allows that a consumer must be informed on the procedure, policies and practices of the forensic analysis that his data is going to be subjected to. |
| RQ12 | The individual participation and access requirement proposes that a consumer must have access to his data throughout the process of investigation. |

| ID | Requirements |
|---|---|
| RQ13 | The accountability requirement promotes recognition that the investigator has to follow the privacy policies set forth for the collection and analysis of the evidence. |
| RQ14 | The information security controls requirement protects collected personal data from unauthorized access, loss, and modification. |
| RQ15 | The compliance requirement incorporates the implementation of an auditing mechanism to ensure that the whole process of investigation complies with the investigative and privacy principles. |

In the openness, transparency, and notice requirement, a consumer must be informed about the procedure, policies and practices of the forensic analysis that his data is going to be subjected to. Similarly, according to the individual participation and access requirement proposed by the researchers, a consumer must have access to his data throughout the process of investigation[66].

The accountability requirement promotes the recognition that the investigator has to follow the privacy policies set forth for the collection and analysis of the evidence. And the information security controls requirement protects collected personal data from unauthorized access, loss, and modification. Finally, the compliance requirement incorporates the implementation of an auditing mechanism to ensure that the whole process of the investigation complies with the investigative and privacy principles.

### 3.5 Mapping the Requirements against Existing Models

The summary of the IoT forensics models evaluated in this section is provided in Table 3.2. In the table, the sign ✓ shows that the requirement is supported by the model and, if the requirement is neither supported nor discussed by the IoT-forensics model, then no sign is given. To evaluate the forensics readiness (RQ1) in an IoT-forensics model, the model is checked for its adoption of a pre-incident preparation approach to IoT-forensics investigations. The model is also checked for whether it provides a means of evaluating an IoT technology for its readiness for forensics; specifically, whether the technology makes use of a privacy-enhancing mechanism in order to not expose consumer private data in cases of investigation.

To evaluate the source of data (RQ2) generated by the IoT ecosystem, the model is assessed by whether it implements data-source triaging methods that would allow an investigator to identify the source of evidence for the case. Additionally, the model is checked for its adoption of some non-repudiation measures in situations where individuals involved in a case cannot deny the integrity of the collected data (RQ3). Taking into consideration the resource-constrained nature of the IoT technologies (RQ4), the automated IoT-forensics tool should not create an additional burden on the smart devices involved in the case. Hence, a model is checked for its employment measures that relieve the computational and storage capacity of the IoT devices.

Table 3. 2 Mapping Requirements against the Model

| Requirements | IoT Forensics Model | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | M1 | M2 | M3 | M4 | M5 | M6 | M7 | M8 |
| RQ1 | | | | | | | ✓ | ✓ |
| RQ2 | | ✓ | ✓ | | | | | |
| RQ3 | | | | | | | ✓ | |
| RQ4 | | | ✓ | ✓ | | | | |
| RQ5 | | | ✓ | ✓ | | | ✓ | |
| RQ6 | | | | | | | ✓ | |
| RQ7 | ✓ | | | | | | ✓ | |
| RQ8 | | | | | | | | |
| RQ9 | | | | | | | | |
| RQ10 | | | | | | | ✓ | |
| RQ11 | | | | | | | ✓ | |
| RQ12 | | | | | | | ✓ | |
| RQ13 | | | | | | | ✓ | |
| RQ14 | | | | | | | ✓ | |
| RQ15 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Considering the privacy of the users of the IoT technologies, a model is assessed for its implementations of means that uphold the consent of the users when collecting evidence from their devices (RQ5). The model is also checked for whether it allows the investigator to inform the user about the reason for the data collection (RQ6). The model is additionally assessed for whether it implements the means of collecting only case-specific data in order to not breach the privacy of other users not involved in the case (RQ7).

The large amount of data produced by the IoT ecosystem is one of the challenges investigators face when investigating an IoT ecosystem. The model is checked for its use of data reduction methods to ease the process of forensic analysis (RQ8). A model must also restrict data collected for a case to be used only in that specific case (RQ9). Principles of digital evidence collection have to be taken into consideration in a model (RQ10). A model must provide the means for clarification of the policies and procedures that govern the owner's data once collected for investigation (RQ11).

When evaluating (RQ12), the model should be assessed for whether it has processes that allow the consumer to have a say on the collected evidence at any time in the process of investigation. A model is also assessed for its implementation measures that can hold the investigator accountable in case he/she perpetrates the invasion of the privacy of the owner of the data (RQ13). Likewise, a model must be evaluated for whether it employs a means of protecting the collected data from access by unauthorized individuals (RQ14). Finally, the model is checked for its inclusion of steps that ensure the compliance by the investigator with the processes and procedures that comply with the principles of the investigation and privacy (RQ15).

### 3.6 Requirements Evaluation

In this section, the requirements identified in the preceding sections are evaluated for usefulness and completeness. The evaluation strategy employed in this chapter is "expert evaluation." The set of the requirements together with the definitions were submitted to three experts to have their views by means of a semi-structured interview. The questions developed for the interview were submitted to the experts using Google forms. Overall, the evaluations by the experts are positive. The questions submitted to the experts included the following.

Q1. Are the elements defined in the requirements useful for user privacy during investigations involving IoT technologies?

Q2. Are there any missing elements in the defined requirements that you think would be important for the protection user privacy in IoT forensics? If yes, please mention them.

Q3. Please indicate on a scale from 1-5 how effective you think each requirement is regarding user-privacy protection for IoT-forensics analysis (1 being the least effective and 5 being the most effective).

The experts involved in the evaluation included three digital forensics experts. They included an associate professor (EX1) and two doctoral researchers (EX2 and EX3). The following contains the feedback given by the experts on the set of the requirements.

EX1: regarding the completeness of the requirements, this expert commented that the requirements are sufficient for obtaining admissible evidence while preserving user privacy. Answering the usefulness question, EX1 indicated that the readiness requirement is a valuable addition that would facilitate the whole process of an investigation. The readiness requirement would help organizations adopt privacy-preserving measures to proactively collect evidence. In addition, the expert said all requirements are useful for an IoT-forensics investigation that upholds user privacy. He also rated the readiness requirements as the most effective, while the other requirements were rated between effective and most effective.

EX2: This expert stated that the requirements are useful and complete and should to be included in models that lead to investigations involving IoT environments. He rated the requirements between 3 and 5; there were none of the requirements that were not effective or not relevant.

EX3: This expert rated two requirements as the most effective (RQ9 and RQ15), while the remaining requirements were labeled as effective. He also acknowledged the usefulness and completeness of the requirements.

## 3.7 Chapter Summary

This chapter highlights state-of-the-art digital-forensics process models specific to an IoT environment that have been proposed in the literature. Given the increased impact of IoT forensics on conventional digital-forensics process models, this chapter identifies the requirements that an IoT-forensic process model should include to be used by IoT-consuming enterprises. The requirements were selected from the literature due to IoT-forensics issues and challenges faced by the digital-forensics community. Subsequently, by evaluating the cloud-forensic process models in the literature for the aforementioned requirements, gaps that need to be resolved in standardizing cloud forensics have been identified and presented here.

# Chapter 4: Model Development of the Privacy-Aware IoT Forensics Model

## 4.1 Introduction

Currently, the best practice for developing digital-forensics models is that existing models are used as a base~~d~~ on which to build new strategies that may be appropriate for the rapidly developing new technologies. Existing digital-forensic models are usually proprietary in that they mainly focus on specific types of investigations. There are no internationally agreed upon standard models, but there are models that have been contributed by some standard-producing organizations, such as NIST and ISO/IEC . From one of our previous works [1][67], it was proven that ISO/IEC models cannot be used for investigating IoT environments, which is to say that existing conventional digital forensics models cannot be used for the development of tools and techniques dedicated to investigations involving IoT. In addition, existing IoT-forensic methods do not fulfill the requirements needed the IoT-forensics models to protect user privacy.

Over the last few years research on IoT forensics has been highly active in the domain of digital forensics, and several works, that are orthogonal to the contribution of this chapter, were introduced as solutions to the challenges in IoT forensics. Apart from these solutions there is and has been a lack of a single IoT-forensic model that combines cloud- and IoT-forensic investigation procedures to support user privacy that is based on a clear set of investigation theories. To address this gap, the authors of this chapter try to answer the question: what are the necessary and essential processes that preserve user privacy in IoT forensics?

To answer this question, the authors have first defined a model for an IoT-forensic investigation. We then propose a model that protects user privacy. In this model, the measures investigators should take for user privacy are considered.

The rest of the chapter is organized as follows: Section 2 discusses the related work~~,~~ and cloud forensics; Section 3 discusses the model development process; Section 4 considers the design and development of the model; in Section 5, the design requirements of the model are proposed; Section 6 discusses the model proposed and its steps; Section 7 adds privacy-protection measures to the model; and, finally, Section 8 concludes the chapter.

## 4.2 Related Work

### 4.2.1 Sources of Evidence in IoT

In recent years, the adoption of IoT devices has increased in private and public organizations. The rapid growth of the use of IoT devices has brought about new frontiers for security and digital forensics issues. For example, IoT devices communicate with each other in a manner that is prone to various types of attacks. Once an incident that involves an IoT device happens, it becomes difficult to investigate. The lack of standards or internationally agreed upon best practices make the adoption of conventional digital-forensics models difficult in the IoT ecosystem. In addition, the heterogeneity of the technologies that a single investigation may involve makes IoT forensics vague. For instance, collecting evidence from the IoT ecosystem may include applications, the Internet, cloud systems, smart devices, and connected systems [3].

In general, IoT-forensics processes usually involve the smart devices, cloud, and Internet. Hence, to conduct digital forensics on the IoT ecosystem, an investigator should be familiar with cloud forensics, network forensics, and device forensics [4, 5]. In each of these three layers, as presented in Figure 1, there are associated issues and challenges that investigators have to deal with to conduct plausible forensics. At the device-forensics stage, sources are identified and evidence is collected from the suspected devices. Network forensics is related to conducting digital investigations on different networks connecting IoT devices to each other via the Internet [6]. Likewise, to relieve the resource-constrained nature of IoT devices, most applications use cloud services for data processing and storage [7]. This makes cloud forensics one of the main parts of IoT forensics. These further exacerbate issues faced by investigators when doing investigations of the IoT ecosystem, especially when it comes down to where to look for evidence in such a complicated ecosystem. Hence, any model developed for the activities of IoT forensics should take into consideration these three layers.

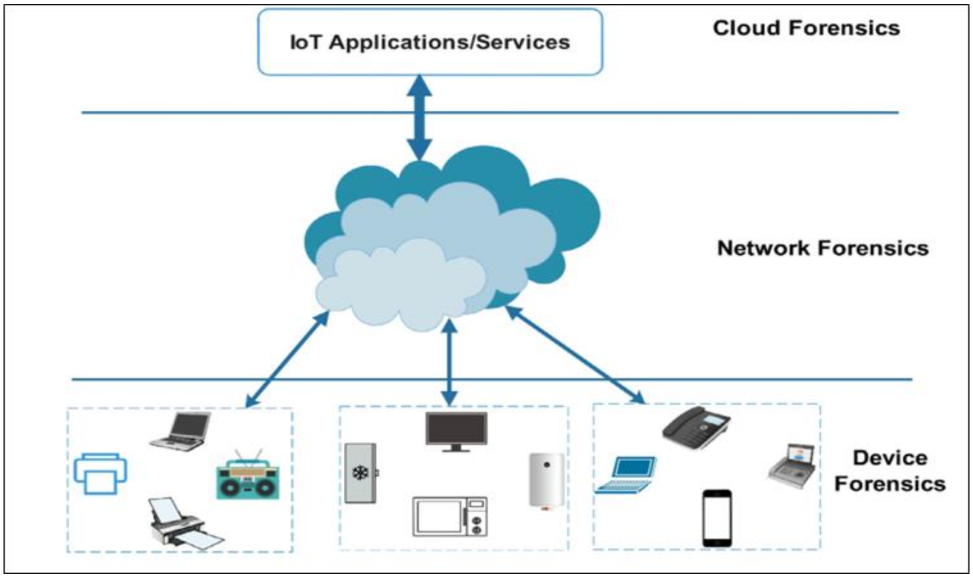

Figure 4. 1 Layers included in IoT forensics

In the IoT ecosystem there are several places to look for the sources of evidence. However, the heterogeneous nature of IoT devices and the huge volume of data to be collected for investigations create~~s~~ issues related to the complexity associated with devices, networks, and cloud services. This may challenge the forensic soundness and the timely manner of collecting evidence. The sources of evidence in the IoT ecosystem can be two core sources, internal and external networks, as shown in Figure 2 [8]. The internal network is related to the local area network (LAN) in which the devices are situated within, while the external network refers to the Internet, web services (API and GUI), and the storage services.

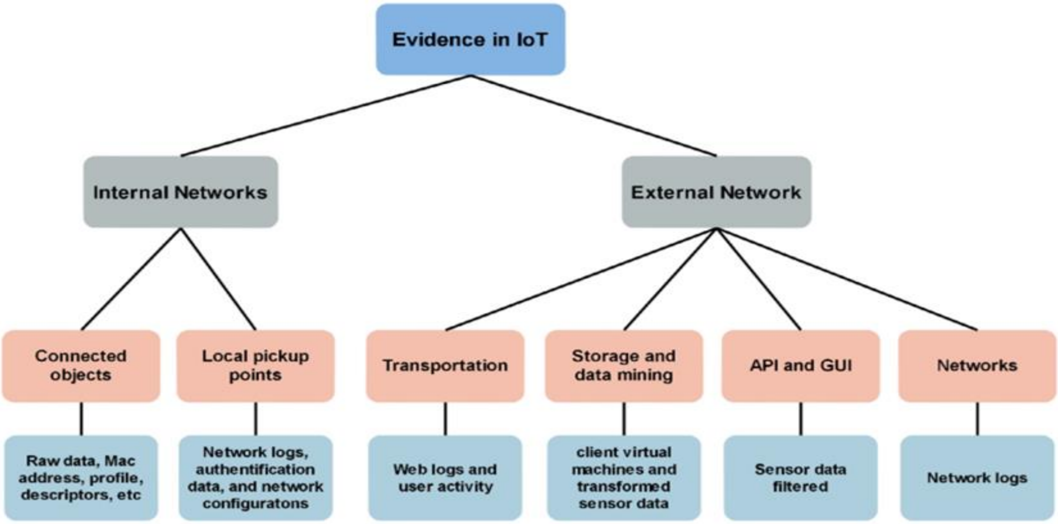
Evidence in IoT
Internal Networks
External Network
Connected objects
Local pickup points
Transportation
Storage and data mining
API and GUI
Networks
Raw data, Mac address, profile, descriptors, etc
Network logs, authentification data, and network configuratons
Web logs and user activity
client virtual machines and transformed sensor data
Sensor data filtered
Network logs

Figure 4.  2 Main evidence sources in IoT [8]

### 4.2.2 Privacy Issues in IoT Forensics

Privacy has been an ongoing issue in the domain of digital forensics [9-11]. One of the main issues reported in the process of conducting investigations in the IoT ecosystem is that the IoT-forensic models proposed in the literature do not uphold or provide user privacy-preserving mechanisms [12, 13]. User privacy is an important issue because IoT devices collect personal and confidential information. Therefore, employment of a privacy-enhancing mechanism is required to protect an innocent user's information when conducting investigations in the IoT ecosystem. Specifically, IoT-forensic investigation models have to employ privacy-protection features [14, 15]. This is because privacy in IoT forensics is not only a technical issue, but a legal and administrative problem, too. Therefore, privacy protection has to be accomplished in the legal and administrative processes proposed in IoT-forensic models [16, 17].

### 4.2.3 IoT Forensics Models

IoT forensics models deal with establishing technical and theoretical procedures that are expected to guarantee the integrity of evidence collected from the IoT ecosystem. Such models may also define fundamental forensic principles needed for the development of tools dedicated to the investigations of IoT ecosystems. For instance, Oriwoh et al. [18] have proposed a model with three investigative zones in which: Zone 1 focuses on the internal network where the IoT devices are located; Zone 2 involves devices at the edge of the network; and Zone 3 is for investigations involving external networks communicating with IoT devices situated within the internal network. Zawoad and Hasan [5] also proposed a model that divides the IoT forensics into device, network, and cloud layers.

Similarly, Perumal et al. [19] proposed a top-down forensic model with a number of steps that help investigators conduct investigations on the basis of the four-tier IoT-reference architecture. The model includes a planning step that consists of authorizing and obtaining the warrant for the investigation. The model subsequently employs a base IoT device-identification step that focuses on finding suspected IoT devices, including the location and the machine-to-machine communication medium used. Once the device and associated communication medium are located, investigators start the third step, which is triage examination. The triage step involves identification of areas with volatile data that may be disrupted easily by investigators. Other steps proposed in the model include common digital forensics procedures, such as chain of custody, lab analysis, result, proof and defense, and archives and storage. Furthermore, Kebande et al. [20] and Sadineni et al. [21] proposed a generic framework and a holistic model, respectively, that consist~~s~~ of the same phases: readiness, initialization, acquisition, and investigative procedures.

In addition, Zia et al. [22] have proposed an application-specific model that involves current best practices in the domain of digital forensics. The model supports the steps, such as collection, examination, analysis, and reporting, from the IoT perspective. Feng et al. [23] also provided an investigative model for the analysis of various cybercrimes in smart city-automated vehicles. The model has four phases that include readiness, deployment, the physical crime scene, and the digital crime scene.

Zulkipli et al. [24] have proposed another a real-time model for IoT forensics. The live investigation model consists of time synchronization, a memory and storage requirement, and a communication requirement. The researchers discussed ~~a~~ pre-investigative measures that include monitoring and automatic collection of data for suspected incidents. Rizal et al. [25] implemented this model in an investigation involving an affected IoT Bluetooth Arduino device using Wireshark software.

Li et al. [26] proposed an IoT forensic model that supports phases such as identification, preservation, analysis, and presentation of evidence in the IoT ecosystem. The researchers implemented the model in the Amazon Echo to showcase the implacability of the model in the IoT ecosystem. Sathwara et al. [27] also proposed a framework that consists of identification, preservation, and analysis. The researchers mapped the forensic activities to four layers of the IoT ecosystem. The sensor layer is mapped to cache and memory analysis of the forensics ecosystem proposed in the study. The network layer is mapped to log analysis, the service layer to the service-level agreement inspection, and the interface layer to fingerprint collection.

### 4.2.4 Evidence Acquisition and Analysis Models

In IoT forensics studies, research includes a focus on modeling the acquisition and subsequent analysis of evidence from the IoT ecosystem. For instance, Servida and Casey [28] discussed ways of extracting evidence from different layers of the IoT ecosystem. The researchers presented a vulnerability assessment based on scenarios that can help investigators conduct IoT-forensic investigations. The study also proposed ways to conduct analysis on evidence from IoT devices. The research reiterated that evidence from the IoT ecosystem can be found on a smart device, the network, and smart applications (cloud). Meffert et al. [29] also proposed a method of acquisition, called Forensic State Acquisition from the IoT (FSAIoT) framework, with a centralized controller that acquires evidence from IoT devices, the cloud, and controllers in the IoT ecosystem. Similarly, Mascarnes et al. [30] have provided a model that helps search pertinent evidence and allows investigators to semantically cluster the data once collected.

### 4.3 Digital-Forensics-Activities Modeling Approaches

The digital-forensic researchers demonstrated their own ways of representing models with different terminologies. Consequently, this has created an ambiguity for readily using these process models in the domain of digital forensics investigations [31]. In response to this ambiguity, several researchers have tried to employ a formal approach that can be used in representing digital-forensic process models. For example, in [32], the researchers proposed a model in a structured approach that feeds a modeling program using digital investigation process language (DIPL), as can be seen in Figure 4.3. As demonstrated in Figure 4.3, the model starts with feeding investigate narrative followed by characterization of representing the narrative into the syntax of DIPL. To convince the research community or add something concrete to the body of knowledge of the domain, researchers are expected to demonstrate their models in real scenarios to show its usability. However, the researchers did not implement this model; hence, its applicability in real cases should be questioned.

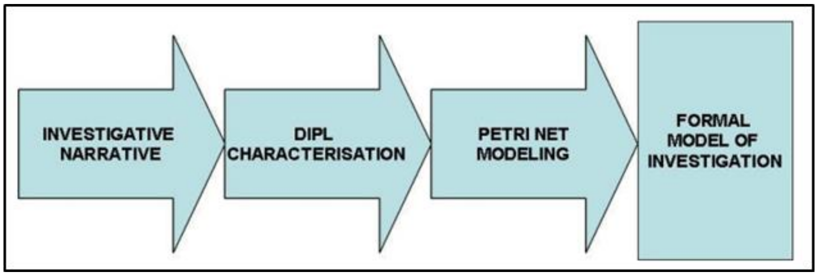

Figure 4. 3 Stages after Stephenson’s model [32]

Toward this same effort, several researchers have utilized unified modeling language (UML) and business process modeling and notation (BPMN) for formally representing the digital-forensic models [33-38]. For instance, Bogen and Dampier suggest the use of UML for structuring and modeling digital-investigation models. The authors employed the use of domain-specific analysis as a tool that can identify, analyze, and document relevant information for an investigation. Likewise, referring to the informal representation of the digital-forensic-process models that exist in the domain, researchers in [34] proposed that UML may help in structuring models. Utilizing UML use case and UML activity diagrams, Kohn et al. (2008?) compared two existing digital-forensic-process models and consequently proposed an activity diagram first and then a use-case diagram that represents the process model proposed by [39], as shown in figures 4.4 and 4.5, respectively.

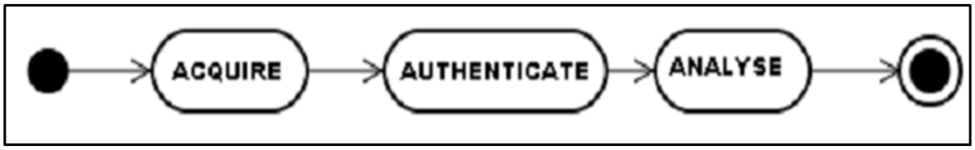

Figure 4. 4 Kruse & Heiser Activity Diagram after [34]

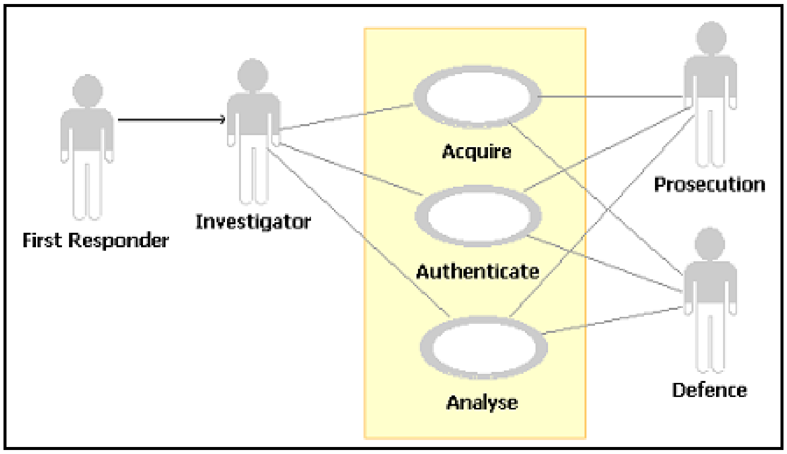

Figure 4. 5 Kruse & Heiser Use Case Diagram after [34]

Regarding representing the digital-forensic-process model proposed by United States Department of Justice, Kohn, Eloff and Olivier (2008) similarly utilized an activity diagram and a use-case diagram, as shown in figures 4.6 and 4.7, respectively.

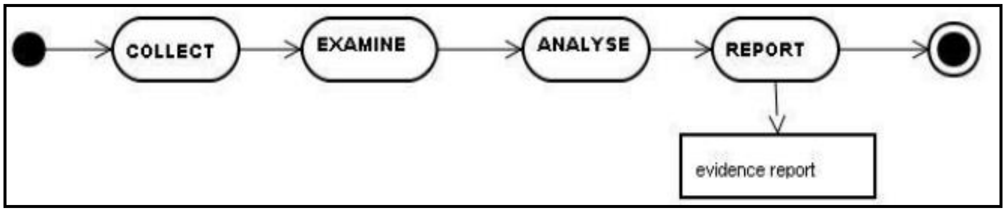

Figure 4. 6 US Doj Activity Diagram after [34]

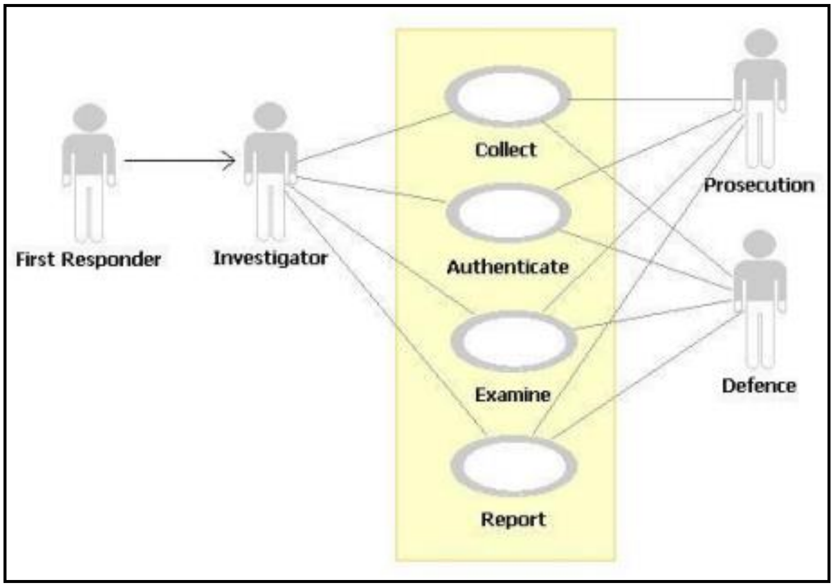

Figure 4. 7 US DoJ Use Case diagram after [34]

The researchers of [35] also utilized UML in representing the most common processes of digital investigations. The UML use-case diagrams the authors employed to represent their work are shown in Figure 4.8, while their UML activity diagram's representation is shown in Figure 4.9.

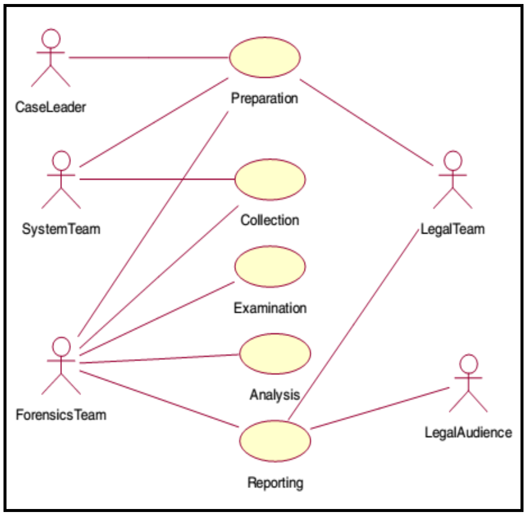

Figure 4. 8 Forensic Process Use Case after [35]

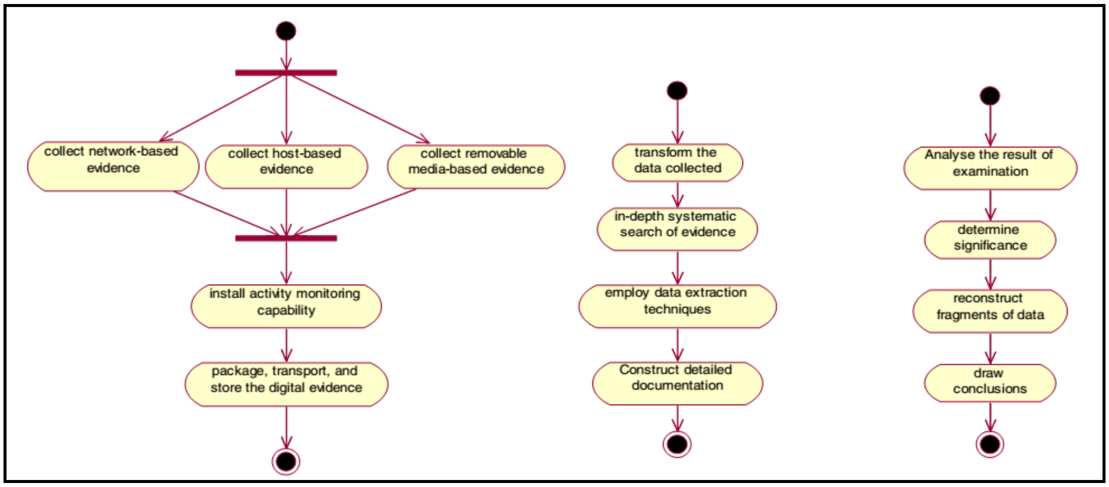
collect network-based evidence
collect host-based evidence
collect removable media-based evidence
install activity monitoring capability
package, transport, and store the digital evidence
transform the data collected
in-depth systematic search of evidence
employ data extraction techniques
Construct detailed documentation
Analyse the result of examination
determine significance
reconstruct fragments of data
draw conclusions

Figure 4. 9 Forensic Process Activity Diagram after [35]

Claiming that existing digital-forensic-process models are in high-level view, Wang and Yu [36] have tried to enhance the model contributed in [33] by representing it with a Petri net. In the representation of the process model with the Petri net, Wang and Yu utilized the terms "action" and "condition" with which they project t-elements and p-elements, respectively, as shown in Figure 4.10.

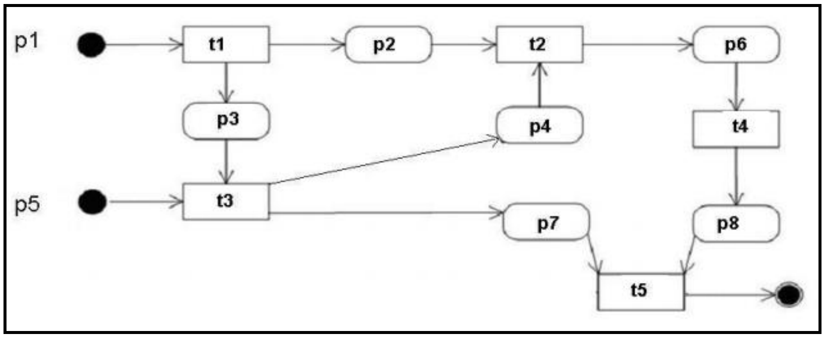

Figure 4. 10 The Perambulation Procedure after [36]

Some researchers have tried to introduce a formal language for digital forensic models. These languages were not as popular as the use of other graphic languages, such as UML and BPMN. Having said that, researchers have shown some agreement on the definition and representation of digital-forensic-process models and an interest has been shown for adopting a formal methodology for the use of process-modeling tools. Following an evaluation of UML and BPMN activity diagrams in the domain of process modeling from three perspectives, which include the capacity of the readily understandable, the adequacy of graphical elements, and mapping to business process execution language (BPEL), it was clear that the latter (BPMN) is more acceptable among users, especially in mapping to the BPEL language [40, 41]. Therefore, in our research the IoT-forensic model would be represented in BPMN activity diagrams, as employed by [37, 38].

## 4.4 Model-Development Process

In the design and development of the model, the design science research (DSR) process proposed by Johannesson and Perjons [42] was followed. The DSR model consists of five main activities that range from problem investigation to requirement definition, through model design and development, and, finally, to the demonstration and evaluation of the model.

The problem investigation activity involves investigating and practically analyzing the problem to be handled by the model. The define requirement activity outlines a solution to the explicated problem in the form of a model; this activity also elicits requirements, which can result in a transformation of the problem into demands on the proposed artifact. The requirements may include both functional and non-functional requirements that the model should fulfill or demonstrate to solve the problem. The design and development activity creates a model that addresses the explicated problem and fulfills the defined requirements. Finally, the demonstration activity uses the developed model in an illustrative case or real-life scenario, thereby proving the feasibility of the model; the evaluation activity determines how well the model fulfills the requirements and to what extent it solves the initial motivating problem. In addition, the process of the model development has been conducted in an iterative way, which is moving back and forth between all the activities.

### 4.4.1 Design and Development of the Model

What led to the development of this model was the lack of a model that could protect user privacy for the IoT-forensic process; the objective was to create such a model. We were searching for a model with a formal description of preserving user privacy that could be adopted by practitioners working in different areas of the IoT ecosystem. The scope of this model is restricted to two of the three perspectives of digital forensics, as provided in [43]. Those are the incident response and law enforcement perspectives of digital forensics. The military perspective was not covered by the model because it is extremely difficult to obtain data on the armed forces and their processes and procedures for conducting digital forensics [31].

The requirements of the model were defined based on the contributions of previous research found in the literature and assertions from digital forensics experts. Based on the defined requirements and the main objective of the research, the authors formulated a set of criteria against which the models in the literature were evaluated. The criteria adopted in this research were based on those proposed in [44] and utilized in [31]. Based on these criteria, the course of model development followed, first, the selection of a group of existing models that would constitute a base for the development of the new model; digital forensics process models are not developed from scratch, but instead use existing models to build new models [8, 45]. The set of models reviewed above in the related work were taken as the validating set.

The criteria included that the model must have practical steps to be taken for investigating a computer, a network, and a cloud. The model must focus on incident response and on the law enforcement agents (LEAs) perspective. As said above, models with a military perspective were not included and removed from the list. The set of models reviewed for adherence to the criteria are shown in Table 4.1.

Table 4. 1 Model reviewed for adherence to the criteria.

| Sources | Model Name | Year |
|---|---|---|
| [46] | Digital Forensic Evidence Examination Processes Model | 2010 |
| [47] | Network Forensic Generic Process Model | 2010 |
| [48] | A Multi-Component View of Digital Forensics Process Model | 2010 |
| [49] | Systematic Digital Forensic Investigation Model | 2011 |
| [50] | "Chain of Digital Evidence" Based Digital Forensic Investigation Process Model | 2011 |
| [51] | Harmonized Digital Forensic Investigation Process Model | 2012 |
| [52] | Towards a SCADA Forensics Architecture Process Model | 2013 |
| [38] | Integrated Digital Forensics Process Model | 2013 |
| [53] | Cloud Storage Forensics Framework | 2013 |
| [54] | A Cloud Network Forensics Framework | 2013 |
| [55] | A Six-Step Collection Process | 2014 |
| [56] | A Heuristic Cloud Forensics Model | 2014 |
| [57] | A Proposed Cloud forensic Approach for Irish Law Enforcement | 2015 |
| [58] | Open Cloud Forensic Process Model | 2015 |
| [59] | A Cloud Forensics Methodology | 2015 |
| [60] | A Cloud Incident Handling And Forensics-By-Design Model | 2016 |

What followed next was the extraction of elements from the base models and then the inclusion of those elements in the new model. The selected elements had to be suitable for investigations in an IoT ecosystem and were short-listed; the ones that were unsuitable were excluded.

In addition, the short-listed elements were given designations in the model in regard to the relationship each element may have had to other elements. A relationship studied at this point builds on the analysis obtained from the literature and on the descriptions the elements have been given in their respective models. When the construction was completed, the model was validated by comparing it to the existing models. Criticisms of previous research in relation to privacy were taken into consideration to gain insights into the design, as well as the pitfalls of implementation, all the while ensuring that the model remains forensically sound.

### 4.4.2 The Design Requirements of the Model

After combining the key contributions from the literature and considering the reviewed models collectively, the model consists of three layers: device forensics, network forensics, and cloud forensics layers. To preserve privacy in IoT forensics, several main requirements must be addressed; these include specifying privacy policies, collecting data only relevant to the suspected case, and using searchable cryptographic techniques to obfuscate sensitive data, such as personal identifiable information (PII). Therefore, these requirements have to be upheld in each of the three layers.

1. At the device forensics layer, the process may involve both live and dead forensics. In the case that devices are sized switched on, it is recommended that the process of investigation starts with live evidence acquisition. The live process must maintain the privacy-protection measures including the privacy policies set forth by the authorities managing the case. Likewise, only data deemed to be relevant to the case are collected from the devices, specifically data related to users potentially afflicted by the case. Finally, before the evidence collected in the live process is transferred for storage, privacy-protecting technique encryption is used to encrypt the collected data in order to deny unauthorized access.
2. At the networks forensics layer, network forensics is usually a live forensic process that includes packet capture and logs collection. There are two types of networks that may be involved in this layer: internal network forensics and external network (Internet) forensics. The internal network is the network that interconnects the IoT devices located in the same premise; the external network is the Internet that remotely connects devices to other devices or owners. The balance between legal enforcement and the privacy interests of users in network forensics requires protecting user privacy at packet collection and analysis. Accordingly, there must be policies and techniques in place with privacy-preserving attributes.
3. At the cloud forensics layer, the cloud services connected to IoT devices are under investigation. In the IoT ecosystem, cloud services are either used for storing or processing the data generated at end devices. In general, cloud forensics involves client- and server-side investigations. However, in the IoT ecosystem, IoT devices are the clients, while the cloud service used by the IoT device is the server side. Hence, the cloud forensics layer is about server-side forensics. Protecting the privacy of an innocent user in the cloud-forensics layer may involve forensic policy generation, collection of suspected user-only data, and

employment of searchable encryption so that investigators cannot meddle with innocent user data.

## 4.5 The Proposed Model

The representation of the model with the three layers is in BPMN. BPMN process diagrams are incorporated within the model to define the flow of the investigation steps in each of the layers. Subsequently, the model is developed by employing BPMN to support instantiation of the model to tailor for different IoT ecosystem scenarios, thereby making the model generic. For the model to assist investigators when applying the concepts and process given in the model, investigators should and must do each layer in conjunction with other layers at a level of detail that permits admissibility of the case at hand.

Each layer of the model is represented in BPMN and discussed separately. The common processes that transverse between layers include abiding to legal terms, evidence preservation, and chain of custody. This is to ensure that all activities associated with the collection of evidence and subsequently the transporting and storing are replicable because there is the potential for the whole process of investigation ordered by the court to be repeated by that court. The model does not require investigators to be specialists in performing IoT forensics, but rather an investigator can easily follow the explicit steps provided in the model. Nevertheless, investigators must observe the requirements that their actions are auditable through clear documentation, be repeatable where possible, using the same tools on the same IoT devices, under the same conditions, be reproducible for the same result with the use of different tools on the same IoT devices, and be justifiable.

The premises where the IoT devices are located are where investigations usually start. The steps proposed in this layer include preparation, evidence source identification and preservation, collection and acquisition, examination and analysis, and reporting and presentation. The privacy-protection protocols employed in each of the steps are discussed accordingly. However, before any privacy measures are added to the steps, the main concepts of the steps are briefly discussed here.

### 4.5.1 Preparation Step

The preparation step is where organizations and individuals using IoT devices proactively prepare and plan before a security incident takes place that may trigger an investigation. What makes this step mandatory in IoT forensics is that an investigation, involving thousands of networked devices that communicate and exchange via the Internet, may easily become exhausting and time consuming. In essence, this preparation and planning step integrates readiness measures into all three layers of the IoT forensics.

IoT devices consist of five modules that include a sensor module, a processing module, an actuation module, a communication module, and an energy module. All the modules participate one way or another in gathering information about the owner, with or without his knowledge. Once owner information is collected, the owner has little to no control over who has access to it. Hence, collecting and preserving evidence in IoT devices has proven to be challenging, which is what makes preparation and planning an essential part of the IoT-device~~s~~ forensics layer. Another main point that makes the preparation step a good candidate for IoT-device-forensics is that the data generated from different IoT devices may not have been stored in a consistent structure or format; this causes conventional forensic tools to fall short of gathering and processing digital evidence from IoT devices.

The preparation step develops strategies so IoT devices are proactively prepared to respond to security incidents when needed. Such preparation is essential to ensure investigative techniques are put in place before an incident involving IoT devices occurs. With this preparation, investigative processes involving IoT devices ought to be optimized by employing proactive steps to guarantee that evidence will be readily available when needed. The preparation strategies that can be taken for improving the investigative capability of IoT devices are discussed in the following paragraphs.

The preparation starts with scenario definition, which is when IoT device owners understand the types of attacks and security risks that come with the use of IoT devices. Owners must also learn the laws and bylaws governing digital forensics that occur~~s~~ in response to attacks involving IoT devices. In addition, with the limited computing and storage nature of IoT devices, owners should focus on evidence-preserving scenarios that would add burden to the already volatile IoT-device resources. Owners might be in a better position to focus on the specifics of attacks rather than trying to get a handle on the wide scope of attack risks. In this way, they will be able to gain insight into typical digital evidence sources that would be useful in an investigation of the particular scenario type.

Even though IoT devices do not store much data, evidence that can be relevant to an investigation may exist on these devices. IoT devices may either be informational or special-purpose devices. For instance, IoT devices such as a smart vehicle’s console and smart watches are informational IoT devices. That is to say, from systems perspective, such devices act as proxies for humans to suggest actions and provide sensors for input. Informational devices can generate and at the same time store data and can be identified as evidential data sources. On the other hand, special-purpose IoT devices have interfaces that provide specific functionality, such as temperature sensors. These devices can have generated data, but the potential evidential data is limited to only those generated data. Therefore, to understand what kind of evidential data can be gathered from IoT devices, owners need to have control over what kind of devices they adopt. This will allow owners to proactively know where to look for evidence in relation to different type of attacks recorded in the scenario generation.

To be prepared for investigations, owners have to understand the data collection requirements associated with their IoT devices. For instance, owners have to know from where to collect evidential data: from the devices and sensors themselves, from a local network that connects devices to each other within the scope of the owner's premises, or from an external network that connects the devices to remote data storage. There are three main requirements that owners should know to ensure the admissibility of the digital evidence from such areas. Time synchronization among devices by using a managed and centralized time service solution is a must. Also recommended is an evidential data storage center with better memory and storage capacity that can swiftly collect and process data generated by the IoT devices.

### 4.5.2 Evidence Source Identification and Preservation

Like conventional digital forensics methodologies, this step is where the investigative procedures start in cases where the preparation step has not been employed. When IoT-forensics practitioners are called in for the first interaction, the source of the evidence is likely to be on the IoT devices. Nevertheless, there are additional activities and tasks that need to be performed that suit the unique characteristics of the IoT devices. In addition, this step also includes identifying whether cloud services were used on the suspected IoT device and identifying cloud service providers relevant to the case. Subsequently, regardless of whether the sources are on the device or on the cloud, investigators need to ensure the proper process of preserving the evidence.

### 4.5.3 Collection and Acquisition

This step involves all three layers of IoT forensics. All other processes that follow depend on the admissibility of the data collected in this step. The following subsections discuss measures that need to be taken throughout the layers of the IoT forensics.

*A. From the devices layer*

This step is concerned with the collection of the IoT devices and acquisition of the data from the sources identified in the devices. This step is mostly based on the state of the devices. In case the devices under investigation are off at the time of seizure, the devices are collected and taken to the lab for investigation. However, if the devices are found on, live forensic acquisition procedures are employed for evidence capturing. The latter is the most usual state of IoT devices and evidence acquisition is deemed challenging due to the transient nature of the IoT device connectivity. Above all, investigators should closely follow existing live forensics best practices to deal with the volatile nature of such evidence.  Since most IoT devices do not have persistent storage, memory capture of IoT devices may help reveal data in the device’s memory. Mobile companion devices used as the IoT management console can also serve as potential evidence [68].

*B. From the cloud layer*

IoT devices use cloud services for data storage and as a point of control. Cloud-layer forensics is the most crucial of the three layers. In collecting evidence from a cloud, it is impossible to physically seize and shut down all the servers in the cloud data center. This is due to the hardware involved and/or due to the fact that other tenants are using the servers and may suffer downtime. In this case, the data collection is almost always live data acquisition that happens via an application programming interface provided by the cloud service provider. This may involve exporting the evidence from the target cloud service while it is still running. It may also involve a capture of the virtual storage and current memory, which is only true in cases using infrastructure as a service (IaaS). In the case of software as a service (SaaS), exporting the actual data stored by the IoT devices would be more likely.

Cloud services used by the IoT devices may contain extra data that may cause a leak of intellectual property or private user data. In addition, the data collection process may suffer from legal issues; for instance, it may happen that the incident has taken place in one jurisdiction and the data resides in another jurisdiction. Hence, investigators have to be aware of legal and privacy-protection laws under which they are acting.

*C. From the network layer*

In IoT forensics, network forensics is a mandatory part of the process. Network traffic can provide investigators with an invaluable source of evidence. IoT devices have unique identifiers such as IP or MAC addresses associated with each device. Hence, network logs are of relevance for IoT-forensics acquisition. IoT networks have become a source of attacks targeting IoT devices [61]. Hence, this step includes capturing network traffic from the internal and external networks to which the suspected IoT device is connected. Since IoT networks generate an enormous volume of data, it is imperative to employ data reduction measures that may help enable near real-time live collection. The data collection and acquisition step is concluded with correlating and consolidating the different evidential data acquired from different sources in the three layers of the IoT ecosystem.

### 4.5.4 Examination and Analysis

This step is concerned with the examination and analysis of the data collected from IoT devices. It is during this analysis when investigators may identify that the IoT devices under investigation have been connected to a cloud service for data storage or processing. The analysis can be conducted online or can be conducted at the laboratory. Data collected from the three layers of the IoT architecture are *examined and analyzed* during this step. Upon analysis of the devices' layer data, if the usage of cloud is identified, this would lead to an iteration of the preceding processes.

### 4.5.5 Reporting and Presentation

This step involves the preparation of reports on the output of the investigation for presentation in a court of law. This step is similar to its counterpart in ~~to~~ the conventional digital forensics process; it does not need major changes. Investigators are required to mind the technical aspects of the investigation to ensure that data contributed by third parties, such as cloud providers, internet service providers, and mobile service providers, is valid [69]. The chain of custody logs must be available to show the stewardship of the data from different levels of an investigation.

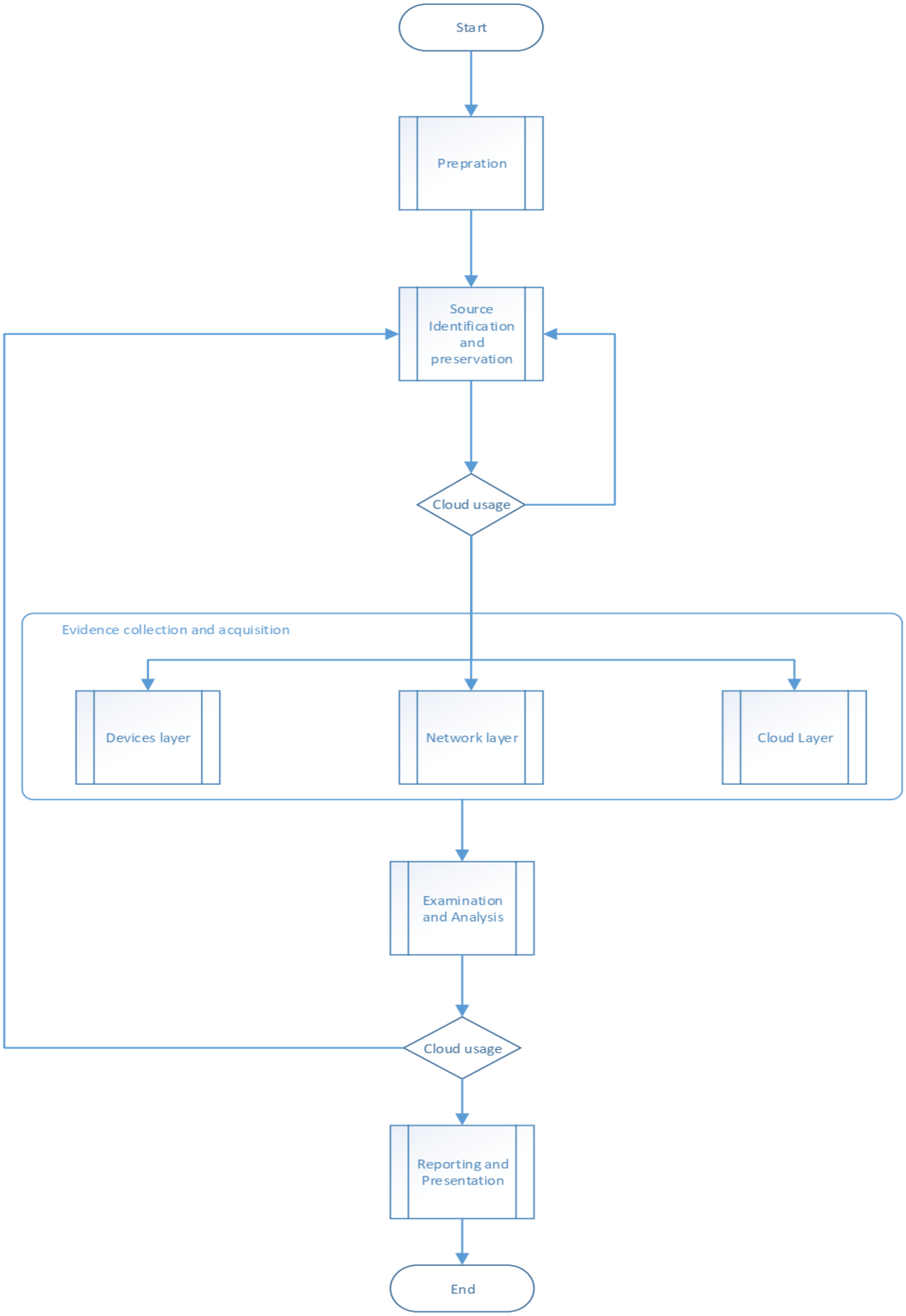

Figure 4. 11 The proposed model

### 4.6 Adding Privacy Protection Measures to the Model

In this section, some privacy-preserving mechanisms are added to the IoT forensics model proposed in the preceding section. Since~~,~~ the model focuses on the business and law enforcement perspectives of digital investigations, privacy measures are discussed in relation~~s~~ to organizations and individuals that use IoT technologies at their locations. Privacy-protection policies and privacy-enhancing technologies deemed necessary in each of the IoT forensics layers are discussed below.

#### 4.6.1 Preparation

The preparation step is particularly important in the two situations: business organizations implement infrastructure and operational preparedness before a security incident takes place to facilitate postmortem investigations; and preparations are made that help law enforcement agents become technically and procedurally ready to conduct forensically sound investigations. However, the main goal of this preparation step remains the collection of information needed for security incidents involving IoT devices when the incidents are happening. Evidential data collected during an incident has to be subjected to some privacy-protecting mechanisms, such as the adoption of authentication and authorization techniques, in order to prevent user privacy issues from postmortem investigations.

Data collected for this aspect of an investigation can be checked for different relevancy and level of privacy needed. For instance, such data can be classified as relevant or non-relevant in relation to the case under investigation, and as private or non-private data in relation to the IoT device user. Based on the level of privacy associated with the data, organizations must include privacy-protection measures in their data protection policies. In other words, the level of trust the data owner has for the investigators should be reflected in the forensics readiness policy of the organization. To that end, data classified as private or confidential should not be accessed–unless the owner permits it. That is only possible when the data owner obtains a privacy-protection guarantee from the investigator[70].

In relation to a crime under investigation, forensic data can be relevant or non-relevant to the case. Relevant data is data considered related only to the crime being investigated. Investigators are expected to minimize data collection to the level of their relevance to the case.

Thus, accessibility to data investigators may have will be based on privacy as well as relevancy associated with an incident. For instance, if the data is labeled as private and relevant, the user has the right to ask the investigator to collect and use the data in a privacy-preserving manner. If the data is private but not relevant to the case, the investigator is not supposed to collect it. Likewise, if the data is neither private nor relevant to the case, the investigator will not need to collect it. On the other hand, if the data is not private but relevant to the case, the investigator should collect it.

### 4.6.2 Privacy preparations needed at the device layer

In this step, as depicted in Figure 4.12, IoT investigators are prepared to acquire only security incident-related data instead of anything that may include personal user information, which may cause a breach of privacy when investigators are gathering evidence. In fact, the device layer contains all the physical resources that collect and control data. Thus, unique approaches for protecting the privacy of individuals and organizations are needed.

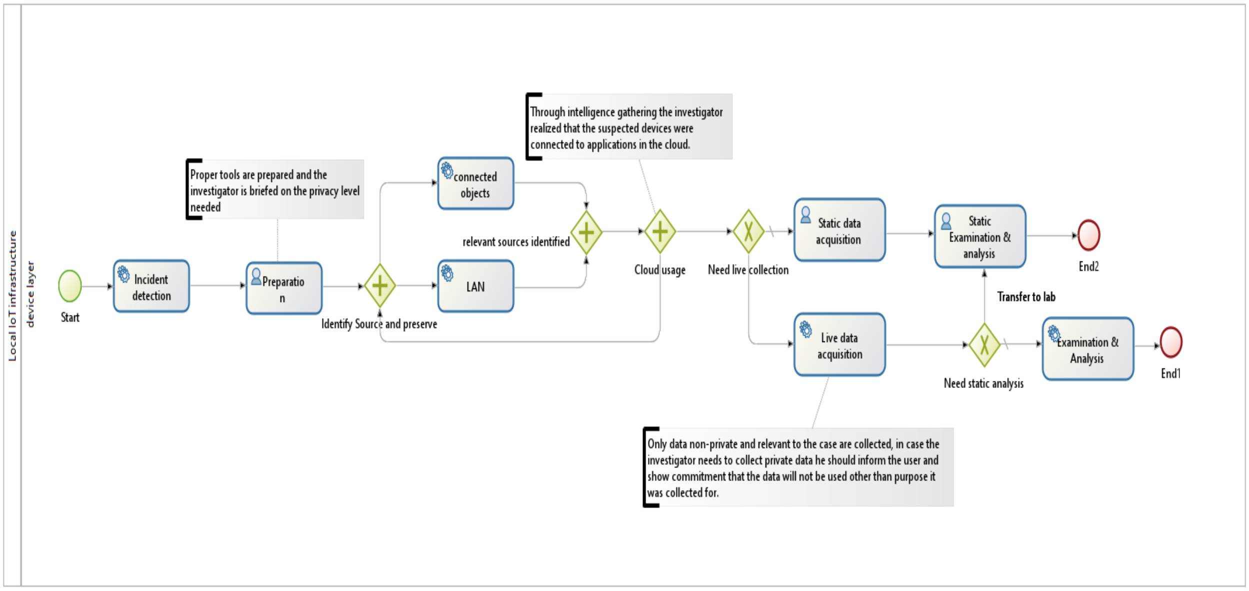
Local IoT infrastructure
device layer
Start
Incident detection
Proper tools are prepared and the investigator is briefed on the privacy level needed
Preparation
Identify Source and preserve
connected objects
LAN
relevant sources identified
Through intelligence gathering the investigator realized that the suspected devices were connected to applications in the cloud.
Cloud usage
Need live collection
Static data acquisition
Static Examination & analysis
End2
Transfer to lab
Live data acquisition
Need static analysis
Examination & Analysis
End1
Only data non-private and relevant to the case are collected, in case the investigator needs to collect private data he should inform the user and show commitment that the data will not be used other than purpose it was collected for.

Figure 4. 12 The Device layer forensics

### 4.6.3 Privacy preparations needed at the network layer

To meet the connectivity demand of IoT technology, network data has increased in terms of scale and complexity. IoT devices send a huge amount of data regardless of the user's will. Hence, mapping between traffic received at some point in the network and an IoT device of origin is challenging. This is the main issue that triggers a privacy concern; this occurs when an investigator seizes and investigates the devices of innocent users, thereby exposing their private or confidential information. Thus, privacy preparation is needed in this network layer and may include providing a mechanism of attribution where network data is linked to IoT devices and their users. The network layer privacy preparation process are shown in Figure 4.13.

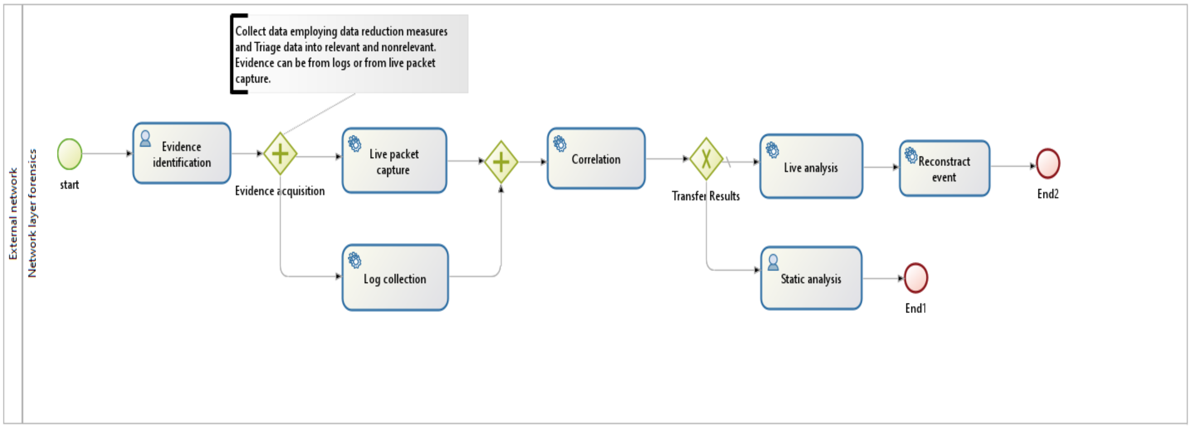

Figure 4. 13 The Network layer forensics

### 4.6.4 Privacy preparations needed at the cloud layer.

Due to the huge volume of data and the distributed nature of the cloud, it is difficult to collect evidence and associate it with a particular IoT device. For example, several devices may use the same cloud instance and it follows that collecting data from that instance might cause a breach of the private data of some other innocent IoT-device users. Privacy preparation is needed in this cloud layer and may include maintaining proper communication channels with the cloud service provider so that investigators may ask the provider to conduct a dedicated search and collect relevant information. Privacy protection activities needed in the cloud layer are shown in Figure 4.14.

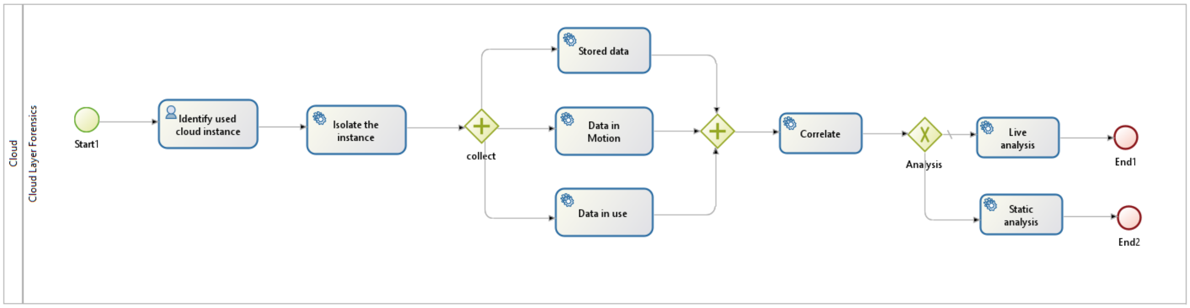

Figure 4. 14 The cloud layer forensics

This step is the beginning of the investigative processes where investigators usually start the work of conducting crime investigations. The complexity of IoT ecosystems makes the process of evidence-source identification and preservation harder. The diverse objects exchanging information among each other complicates the identifications of sources of evidence in the IoT ecosystem. Deployment of user privacy-preserving mechanisms associated with this step is influenced by this heterogeneity as well. In this case, it is essential to adopt user privacy-preserving policies that allow investigators to differentiate sources with sensitive data from those with less sensitive data.

## 4.7 Evaluating the Model

This section tests whether the model is both theoretically and practically valid to answer to the lack of a standard and whether it can be a valid solution for the issues of privacy for IoT forensics. The design science research model adopted to lead this research believes that the goals of an evaluation can be summed up into two, to conduct *a formative evaluation* to assess while the model is still under design and development in order to obtain information about how to improve it during subsequent design activities, and to conduct *a summative evaluation* to assess a model when it has already been finally designed and developed. The research community stressed on the importance of evaluation, providing a number of evaluation goals, strategies and methods. The evaluation strategy provide in the research model employed in this research can in one hand be an ex-ant or and ex-post, and artificial or naturalistic on the other hand. Ex-ant evaluates the model without being fully developed, while ex-post evaluates the model by implementing it in an experiment. With the naturalistic evaluation real forensics practitioners use the model in real-world practices, while with the artificial evaluation the model is used in simulated environment. In this research, the naturalistic evaluation where a group of practitioners and experts have evaluated with based on their experience is conducted.

The evaluation incorporates the views of experts and practitioners, which was necessary because it drew upon knowledge that could not have been obtained otherwise through the analysis of the literature. In conducting the evaluations, there were methods set forth in previous research; the current authors chose those that met our particular research needs. This evaluation employed the interview evaluation as has been used in a number of digital forensics research chapter for validation of models [71]. The model design was conducted on the basis of previous research, but this might be considered as self-reflection. In order to remove any subjectivity from the design process, the evaluation was conducted by having an interview in a semi-structured manner.

### 4.7.1 Identifying and nomination of experts

It was not easy to identify and select experts who could evaluate the model because it is difficult to find experts who extensively understand investigations involving IoT ecosystems. The experts who participated in this evaluation were selected on the basis of their length of time teaching and experience in the domain of digital forensics. Fortunately, three digital forensics experts volunteered to participate in the evaluation. The experts included an associate professor (Expert 1), and two doctoral researchers (experts 2 and 3). The feedback from the evaluation is provided below.

A presentation about the approaches and steps provided in the model was submitted to the evaluators. The key criteria used for this research to assess the model follow.

### 4.7.2 Utility of the Model

Q1. In reviewing the model, please identify any aspects that would not be represented in the process as carried out when investigating IoT ecosystems.

Q2. In reviewing the model, please identify any important aspect of IoT forensics that would protect user privacy and are not covered.

Q3. Please identify any aspect of the model that you feel could be improved to protect user privacy.

### 4.7.3 Usability of the Model

Q4. Please rate this model in terms of usability from 1 to 10 (with 1 being very difficult to use and adopt and 10 being very easy to use and adopt).

### 4.7.4 Feedback from Participants

The BPMN representation of the model together with some prescriptive knowledge about the steps were submitted to the evaluators. In general, a process model can be evaluated based on three aspects, such as whether the model is theoretically valid, whether the model is usable, and whether the model provides explanatory or prescriptive knowledge for its users. A model can be regarded as theoretically valid if to some degree it adheres to the principles upon which the process is developed. The model can be usable if it can be used in its intended environments. Likewise, the model has prescriptive power if it steers the process of investigation and recommends some courses of action.

However, in this research, the model was evaluated for its utility and usability. In assessing the utility of the model, it was evaluated for its usefulness, functionality, and fitness for the purpose of privacy protection. Experts were asked if the model was suitable for describing a privacy-protecting IoT-forensics process. Likewise, for usability, the experts were asked to rate the model for its ease of use and adoption. In general, the feedback from the experts was positive. However, there are some comments that need to be addressed in the model. The feedback from each of the experts follows.

**Expert 1:** commenting on the usefulness the model, the expert said that, with the inclusion of the readiness part in the model, “organizations will make optimal use of evidence in a limited period of time with minimal cost.” However, the expert proposed the need for a technique that reduces the large volume of data that may contain private data of innocent users. The point raised by the expert is very important and has been accommodated in the model by promoting the collection of data relevant to the case. The expert also proposed the need for a process that protects cloud co-tenant data from un-consented data collection. This step can be achieved by isolating the suspected cloud instance, as is currently provided for in the model. Regarding the usability of the model, the expert commented that the model is easy to use for handling IoT forensics.

**Expert 2:** this expert did not highlight any missing or extra steps in the model. The expert commented that the model provides adequate guidelines for investigators to understand and conduct appropriate IoT forensics. Expert 2 rated the model as an easy to use model.

**Expert 3**: the expert proposed the inclusion of a "correlation" step that compares and fuses different evidential data collected from different parts of the IoT ecosystem. In addition, Expert 3 discussed the importance of restoration once needed data is collected. The model was rated as easy to use and adopt as well.

To optimize the result on the system, a systematic and regulated approach to testing was conducted by seasoned specialists and those associated in the field of computer forensics, antimalware practices, information security and other related areas. The output of the analysis based on the usability and utility is shown in Table 4.1.

Table 4. 2 Result of the evaluation

| Factories | Usability | Utility |
|---|---|---|
| Positive | 90% | 85% |
| Negative | 10% | 15% |

### 4.7.5 Chapter Summary

In order to develop the model, the authors have conducted a detailed study on the issues and challenges pertaining to IoT forensics. It was subsequently deduced that the main issue that has not yet been covered by previous research is the protection of user privacy when investigating IoT environments. Consequently, requirements that have to be fulfilled by the model to answer to privacy-protection issues have been identified. Subsequently, a number of conventional digital forensic models were reviewed for privacy-protection possibilities. As a result, the model proposed in this chapter has achieved success.

# Chapter 5: Using Searchable Encryption for Privacy Protection of Proactively Stored IoT Network Data

## 5.1 Introduction

In this chapter, a network data encryption technique is employed that preserves user privacy. The technique is a solution to the privacy issues faced when investigating incidents in the IoT ecosystem. There are a number of tools used by network administrators for network data collections [1]. The process of recording network data before an incident is a pre-investigative forensics procedure [2]. Recording the network data flow is also used for identifying network users and the types of applications used on the network [3]. Network flow recording may reveal what websites were visited, when visited, and how much data the user transferred out of or into the network. Thus, such information is a serious threat to Internet and web privacy [4]. For instance, if the recorded data is compromised, it can reveal a great deal about the users. Accessing the collected network data for forensics is another way of compromising user privacy as well [5].

Investigators usually use network-analyzer tools to read network data; these include Onion NetFlow Traffic Analyzer (NTA) [6], which enables the capture of data from networks and the conversion of it into charts and tables that can be easily interpreted. NetFlow analyzer is another tool that uses Cisco NetwFlow [7] for bandwidth monitoring and information gathering on network users, applications, and other features. Since~~,~~ all these tools and the collected network data may expose private user information, there is a need for mechanisms that allow investigators to analyze the data while protecting user privacy. This chapter discusses techniques that were proposed in this research to aid in protecting user privacy. Subsequently, the proposed techniques were employed in a case study that involved a simulated IoT-user company that uses a forensic readiness procedure to collect network data, store it in the cloud, and retain it for a long period of time. Analyzing such data for evidence may constitute threats to user privacy, as the data may contain user-specific information that may jeopardize owner privacy associated with human rights. A searchable encryption-based solution is proposed to address the privacy threats targeting recorded data. The data is considered to be stored in a cloud server. The encryption prevents a cloud service provider from recognizing user-sensitive information stored in the data. In addition, with the use of searchable encryption, investigators can access parts of the data that are relevant to the case. The use of searchable encryption also helps reduce the amount of data by allowing investigators to access only those deemed important to the case.

## 5.2 Background

This section briefly introduces network data collection and searchable encryption methods. Network packets, flows, and logs are the main data collected for forensics investigations. Subsequently, the collected network data is stored for future forensics analysis. The stored data is prone to manipulation by the potential alteration or deletion of some convincing evidence. In addition, at the analysis stage private user information may be exposed illegally to third parties. Researchers use encryption to secure and protect private information. In the following, we briefly introduce three basic types of network data-collection methods and the searchable encryption process most used for privacy protection.

### 5.2.1 Network data collection

In networks (both wired and wireless), packets are a significant carrier of effective information. In the IoT context, IoT devices send packets that include user, sender-device, and destination-device information. Packets have various formats based on the different protocols used in the network [8]. Normally, packets consist of a header part and a payload part. The header guides the packet to transmit within a network and mark the source information of the packet [9]. In the data collection, the header is used for the identification and filtering of the packets. For instance, header-based data-collection methods classify the packets into several flows based on their IP addresses, port numbers, and the protocols contained in the header. The payload contains the data transferred or communicated in the network [10]. It is the data contained in payload that needs to be encrypted for user privacy protection [11]. Popular tools used for network packet capture may include, but are not limited to, WireShark, Tcpdump, and Ettercap. Capturing packets is a traditional network data-collection method. It is the most commonly used method to conduct network data collection [12-14]. However, with the widespread use of IoT devices, smart mobiles[72], and clouds, the increased speed and volume of networks have overwhelmed the use of packet-capture methods for forensics [1, 10]. As a result, researchers have abandoned the idea of capturing all packets and opted to employ packet-sampling mechanisms [15, 16].

Network-flow collection is another important network data-collection mechanism [17]. Network flow is a set of packets that contain the same characteristics passing through a specific observation point over a period of time. In most cases, flow collection happens on the network core devices. Flow can be collected at network edge nodes and on hosts, as well. However, edge nodes and hosts can only record inbound and outbound network flows that pass through [18]. The Cisco NetFlow is the most common network-flow recording protocol. NetFlow is a format and tool that implements a five-tupleflow data-collection mechanism: source and destination IP addresses, source and destination ports, and protocol types [19].

Network logs are also used for network data collection for network event~~s~~ reconstruction [20]. Event logs record user traces and event status. Log files are stored in persistent storages. Log files usually occupy a large portion of the memory, with low information density, and apply complex file formats, which makes manually handling of log files difficult.  Hence, automatic solutions have been proposed to counter the problem [21].

### 5.2.2 Searchable Encryption

Privacy-preserving searchable encryption is used in many domains, including: healthcare systems to protect patient-related information [22]; digital forensics to collect case-related information and to protect data from destruction or alteration [23]; email-routing systems to protect emails from adversaries [24]; secure audit logging to protect from adversaries who try to tamper with the contents of the audit logs [25]; and financial institutions to fulfill the privacy requirements of customers [26].

There are normally two types of searchable encryptions, searchable symmetric encryption (SSE) [27] and public key encryption with keyword search (PEKS) [19]. For SSE the data owner uses the same key for both encrypting and decrypting. The primary aim of the SSE is to enable data access only by the data owner. To enable data access by other users in SSE, the data owner has to provide them with the secret key of the encrypted queries. SSE is computationally more efficient compared to PEKS. On the other hand, when using PEKS the encrypted data is shared among multiple users where private and public keys are used [28].

The searchable encryption process and the steps needed by data owners and other stakeholders to conduct a search on encrypted data include; the data owners first generate plaintext indexes for the data, using either forward-indexing [29] or inverted-indexing [30]. The two types of indexing can be represented by various types of data structures. This results from different indexing methodologies, such as tree-based [31], bloom filters [32], bucketization indexing [33], etc. The indexing techniques are aimed at performing search operations that are either linear or sub-linear to the total of the data stored. The indexes themselves are encrypted with normal encryption schemes including RSA [34], RSA [35], functional encryption [36], deterministic encryption [37], predicate encryption [38], etc. once the indexes are encrypted, the owner subsequently needs to encrypt the plaintext data using either public key or secret key encryptions. Finally, the data owner should send both the encrypted data and their corresponding indexes to a storage server in the cloud or to local premises [39].

In the search process, to return the documents in an order of decreasing relevance via given trapdoors, various keyword weight measures, such as term-frequency (TF), or term-frequency-inverse document frequency (TF-IDF) [40], etc., are included in the searchable indexes to determine information in a ranking order. To calculate the similarity measures there are several techniques used, including cosine similarity [41], Jaccard coefficient [42], locality sensitive hashing [43], and inner product operation [44], which calculate the rank of the data based on a given query. The ranking information, returned in response to the search query, is also encrypted using order preserving encryption (OPE) [45], Paillier encryption [46], or fully homomorphic encryption [47][73] schemes. Such schemes preserve the order of plaintext-ranking information and enable ranking based on the encryption-ranking information [48]. The schemes must be carefully selected based on the confidentiality or sensitivity of the data because the schemes have their own advantages and disadvantages with respect to privacy, time, and precision [49].

In the end, the data owner should, therefore, provide other stakeholders with trapdoors, i.e., encrypted queries along with a value 'k' on the server, which returns the top-k data with the help of the ranking information available in the indexes. Returning the top-k relevant data reduces the network traffic and also delivers the information required by other stakeholders, one of whom may happen to be an investigator [50].

### 5.3 Related Work

Network forensics is about capturing, recording, and analyzing network-flow data for event reconstruction in order to identify what went wrong with a network system behaving abnormally [9]. Network forensics can be viable only in environments where network security policies, such as authentication, firewall, and monitoring, are enforced. Most network security systems have the ability to collect data before or during incident detection [9]. Nevertheless, such recorded network data may raise issues of privacy because the data may contain user-sensitive information [51]. In addition, data collected for this purpose is stored most of the time~~s~~ as plaintext either in a dedicated server at a local infrastructure or in a cloud instance. Thus, in practice, encryption technologies are used to protect the privacy of confidential user data [52]. However, encryption may prevent investigators from conducting a proper investigation on encrypted data. For instance, making use of RSA to encrypt stored data and further storing the encryptions on a remote server may make search-over data somewhat impossible on the server side [53].

Searchable encryption has been introduced to tackle the dilemma that arises from the fully encrypted data [26, 54]. The idea behind the searchable encryption is that the investigator sends a search token to the server so that the server fulfills the search over encrypted data without accessing underlying data and query content. As raised previously, searchable encryption strategies are of two main types; symmetric searchable encryption (SSE) [55] and public key encryption with keyword search (PEKS) [56].

A searchable encryption scheme was used in [57] to allow the forensics investigator to query for data matching some specific keywords in the context of email servers. First, the disk image is analyzed and an index file that matches keywords to files (or sectors in the disk image) is generated. The index file and the image are encrypted at this point. Then, the forensics investigator generates a list of keywords, which are relevant to the investigation, and passes the list to the data owner. The data owner can then generate a trapdoor, which is a data structure that allows the investigator to search the encrypted index file for a given keyword or a set of keywords. The investigator can then ask the data owner for the specific location of the file on the disk interactively. A notable limitation of this scheme is that the data owner can potentially hide information from the investigator since he/she is in charge of creating the index file, the trapdoor, and decrypting the files. Another limitation is that the data owner gains sensitive information about the case from the keywords provided by the investigator to obtain the trapdoor.

The solution proposed in [58] aimed to prevent the server administrator from learning what the investigator is looking for. To that end, the investigator generates a public/private key pair and shares the public key with the administrator. Then, the administrator divides the documents into keywords and encrypts them with the public key provided. The investigator encrypts its ‘n’ keywords with the same public key and transforms them into a polynomial of the degree ‘n’. After that, the investigator sends the administrator the coefficients of that polynomial to hide the actual encrypted keywords. Finally, the server administrator makes a similar transformation of the files into coefficients and using the coefficients from the investigator, the administrator can determine which files contained keywords of interest to the investigator. An important point that is not sufficiently discussed in the chapter is how the administrator can prevent the investigator from asking for keywords that are irrelevant to the investigation. Also, it seems that if the number of keywords of interest to the investigator is small, the administrator can more easily brute-force the coefficients and obtain the keywords. This is in contradiction with user-privacy preservation since the investigator should ask only for the minimum amount of information that allows him or her to close the case.

Finally, [59] provided a searchable encryption scheme with the following features: (a) the keyword search is non-interactive, meaning that the forensic investigator does not need to contact the data owner every time he/she wants to issue a query, and (b) the data owner remains oblivious to the queries (keywords) of interest to the investigator. The data owner determines which are the keywords that can be queried for and establish a threshold of the keywords that need to be present in the file for disclosure. Very basic keywords, such as pronouns, can be excluded to prevent trivial attacks. The approach is based on Shamirs secret-sharing scheme; that is, the key for decrypting a file is constructed based on the keywords in that file and that of such keywords reveal(s) the key. The authors acknowledge two main limitations to the scheme: first, the need to have exact keyword matches for searching; and, second, there is the possibility that the investigator performs brute-force/dictionary attacks on keywords to reveal the key. In addition to these limitations, there is the problem of a potentially malicious data owner who wants to limit the ability of the investigator retrieving data. This would be as simple as blacklisting some keywords.

## 5.4 Problem formulation

Communication between smart devices on the same local premises with each other or with external devices in the IoT ecosystem is facilitated by different network protocols. IoT-using organizations use network-monitoring tools to identify and isolate malfunctioning systems. Such monitoring tools may include IDSs or IPSs. Monitoring tools usually employ some kind of network-flow recording so that the recorded data can be used in future investigations. Securing the collected network data is one of the most cited issues. In general, protecting personal information in collected network data can be accomplished by using forward and backward privacy. Forward privacy ensures that network data collected before a point of compromise remains secure even if the encryption key is compromised, which is to say that an attacker is not able to search any newly updated data using previous trapdoors. Likewise, within any period that two trapdoors on the same keyword happened, backward privacy ensures that it does not leak information about the files that have been previously added and later deleted [60].

Usually, organizations employ the latter for the protection of user privacy in network-flow recoding. For instance, organizational network security policies may include the retention of network flow and clickstreams for a period of weeks, months, and even more [61]. It is this stored data that must be protected from unauthorized access so that exposure of private data is limited only to the time of breach of policy or network compromise persists undetected and is no longer a danger/threat. Thus, the focus of this chapter is the use of backward security for network data collected for forensic readiness. This chapter thus proposes the use of an encryption that enforces the privacy policy, but one that can be searched in case an investigation is required for incidents involving a compromised IoT-network layer. Using searchable encryption provides two important properties in this regard. First, data is encrypted and stored in a cloud server for future investigations without the cloud service provider knowing either the public key used by the owner organization to encrypt the data or the corresponding private key meant for decrypting the data. Second is mitigation for a situation when the investigator and the cloud service provider do not trust each other. To prevent the cloud service provider (who may play with the data for reasons of his/her reputation) from learning the investigation subject, the investigator will provide the cloud service provider keywords that are in an encrypted form. Likewise, to prevent the investigator from retrieving irrelevant data, the cloud service provider will later check whether the investigator cheated for obtaining information of other co-tenants from the server. In this situation, to distinguish the relevant data from the irrelevant data, the data involving the specified keywords are considered relevant data and those not involving the specified keywords are irrelevant. Then, the investigator is restricted to performing a search on data that involves the specified keywords. In short, the keywords specified by the investigator are presented as trapdoors by the data owner, and therefore the service provider cannot learn anything about the keywords and the investigator cannot learn more than the search results.

### 5.5 Proposed Scheme

In this section, the proposed PEKS model (Figure 1) meant to protect the privacy of the forensic data for IoT ecosystems is presented. The model has three agents: a data owner, who is the IoT device user; an investigator, who is called in to investigate incidents involving the IoT device owner's infrastructure and the corresponding server where the data is stored; and a cloud service provider, who is the owner of the cloud server where the forensic readiness data is stored[74].

- Data owner: the network data is collected, keyword indexes are created for the collected data, the data is encrypted, and uploaded. The data owner is also responsible of creating trapdoors for the investigator.

- Investigator: initiates the search by asking for trapdoors for the keywords that need to be searched for the case. Once the authorized trapdoors are received, the investigator will conduct the search to retrieve documents containing the keywords. The returned data is then used as evidence.

- Cloud service provider (CSP): CPS's server is responsible for storing the encrypted data and the corresponding indexes and responding to queries from the investigator.

The formal definition of the model is given in the following. The model consists of the following algorithms*: Setup, KGen, PEKS, AuthToken, Trapdoor, and Search.*

- *Setup(m):* takes as input a secure parameter $m$, and outputs system global parameter $\sum$.

- *$KGen(\sum)$:* takes as input system parameter $\sum$, and outputs public/private key pairs *$(PK_s, SK_s)$, $(PK_r, SK_r)$* for the data owner and investigator, respectively.

- *$PEKS(PK_r, SK_s, I_w, w)$:* takes as input the public key of the investigator $PK_r$, the private key of the data owner $SK_s$, and the index set $I_w$ of keyword $w$, and outputs ciphertext $CI_w$ of the index set $I_w$.

- *$AuthToken(PK_r, w)$:* takes as input the public key of the investigator $PK_r$, and the authorized keyword $w$, and outputs an authorization token $A_w$ of keyword $w$.

- *$Trapdoor(PK_s, SK_r, A_w, w)$:* takes as input the public key of the data owner $PK_s$, the private key of the investigator $SK_r$, the authorization token $A_w$ of keyword $w$, and outputs the trapdoor $T_w$.

- *$Search(CD, T_w)$:* takes as input the ciphertext data $CD$ and a trapdoor $T_w$, and outputs ciphertext $CI_w$ of index set $I_w$.

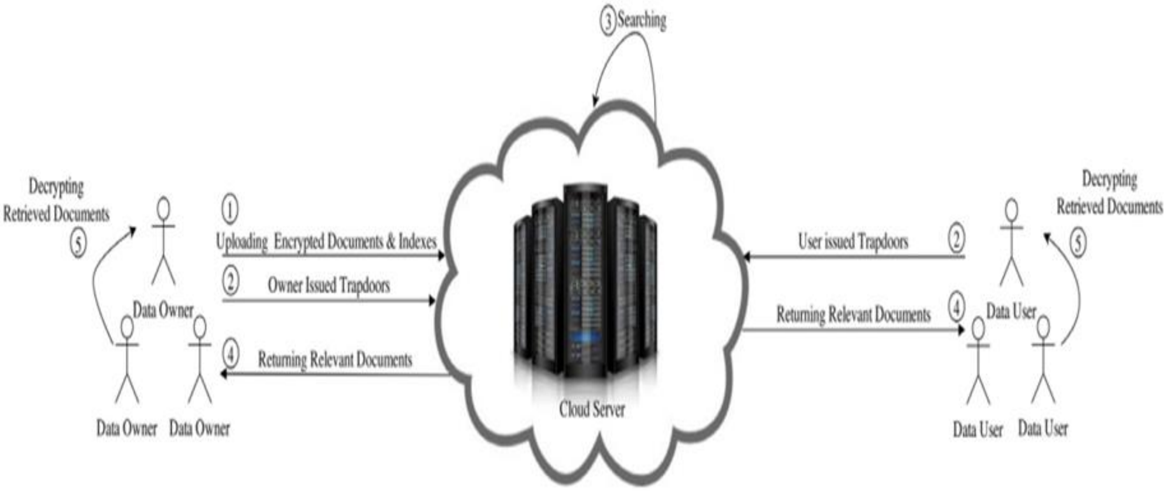

Figure 5. 1 Proposed model

There are a number of privacy requirements that the model should fulfill. These requirements include the following points.

*Data privacy*: the content of the encrypted forensics data uploaded to the cloud server should not be revealed to the cloud service provider. Likewise, the cloud service provider must not be able to deduce keywords corresponding to the investigator's authorized trapdoors.

*Index privacy:* the indexes contain IDs and corresponding encrypted keywords; hence, the cloud service provider should distinguish the indexed keywords corresponding to different documents.

*Trapdoor privacy*: to achieve trapdoor privacy, the cloud provider should not distinguish the trapdoor of one query (submitted by the investigator) from the trapdoor of another query in a polynomial time.

## 5.6 Implementation

The implementation began with network data collection followed by the process of encryption using the proposed model to protect privacy during investigative procedures.

### 5.6.1 Data collection

The data collection was conducted at the internal and external networks the IoT devices communicate with each other and with a remote cloud for backup. The cloud server was prepared with ownCloud installed on an Ubuntu server [62]. The IoT device consisted of three nodes, two of which were located on the same internal network while the third was remotely accessed via the Internet. All three nodes used ownCloud as a backup server, using three different user profiles on the server. The three IoT devices were simulated by using the Android Things platform [63]. The data collection was achieved by using fprobe to record the traffic that passed through tapped network interfaces. Then nfcap was used for the collection of the recorded data in NetFlow format. Each file contained five minutes of recorded data. Once the data collection was completed, the process of encryption and the uploading of the files to the ownCloud started, as follows.

### 5.6.2 Data Encryption Process

The data owner neither trust the cloud service provider nor the investigator. As a result, the data owner encrypts the collected data and uploads indexes and the encrypted data to the ownCloud server. This happens before any incident has been detected. That is to say, the data owner is proactively collecting and encrypting the data. Using the model the data owner then performs the following steps. A low fidelity prototype, which consists of three modules, build-index, Trapdoor, and search, has been developed using Python (Pycryptodome) for the implementation of the searchable encryption scheme. The code lines of the three modules can be found at Appendix B.

***Setup(m)*:** *takes as input the secure parameter m:*

1. *The data owner chooses an additive group $\mathbb{G}_1$ with a large prime order q, and a generator P.*
2. *The data owner selects a trapdoor permutation function (TPF) $f: x \to y$, where X and Y are two sets of strings with length l. Then, the data owner executes $Gen(1^m)$ to generate $f$'s public/private key pair (pk, sk).*

3. *The data owner subsequently~~,~~ selects three resistant-collision hash functions:* $h_1 : \{0,1\}^f \rightarrow \mathbb{Z}_q$, $h_2 : \mathbb{Z}_q^* \times \{0,1\}^f \rightarrow \{0,1\}^*$ and $h_3 : \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \{0,1\}^{(l+logq)}$.

4. *The data owner constructs an SL* ***List,*** *which is the state? Stated? list of the collected data. Then, the data owner initializes* ***List****[w].c = 0 and* ***List****[w].*$st_c = st_{0wi}$, *where* $st_{0wi}$ *is a random string of length* $l$ *and* $w = w_1, \dots, w_{|w|}$.

5. *The data owner sets system parameters* $\sum = \{pp, sp\}$. *Then, the data owner publishes parameter* $pp = (q, P, \mathbb{G}_1, h_1, h_2, h_3, f, pk)$ *and keeps* $sp = (sk,$ ***List***$)$ *secret~~e~~.*

The key generation process is accomplished by implementing the following steps.

***KGen*** $(\sum)$ *takes as input the system parameter* $\sum$, *and performs the following:*

1. *Randomly chooses* $SKs = s$, $SKr = \alpha$ *as respective private keys of the data owner and the investigator.*

2. *Computes* $PKs = s \cdot P \in \mathbb{G}_l$, $PKr = \alpha \cdot P \in \mathbb{G}_l$ *as respective public keys of the data owner and the investigator.*

The encryption of the collected data and the corresponding indexes are achieved with the following steps.

***PEKS*** $(PK_r, SK_s, I_{wi}\ w)$ *takes as input the parameters* $PK_r = \alpha \cdot P$, $SK_s = s$ *and an index set* $I_w = \{ind_1, ind_2, \dots, ind_n\}$ *of w, and performs the following:*

1. *The data owner computes* $K = s \cdot PK_r \in \mathbb{G}_1$.

2. *If* $n = 1$, *then the data owner selects a random index* $ind^*$, *and sets* $I^*_w = \{ind_1, ind_2, \dots ,ind_n\}$ *and* $n = 2$, *where* $ind_2 = ind^*$. *Otherwise, sets* $I^*_w = \{ind_1, ind_2, \dots ,ind_n\}$.

3. *The data owner encrypts indexes as follows:*

    - *Searches* ***List*** *to obtain* ***List****[w] and sets* $j =$ ***List****[w].c and* $st_{jw} =$ ***List****[w].*$st_c$.

    - *For each index* $ind_m$, *the data owner computes* $st_{(j+m)w} = f_{sk}(st_{(j+m)w})$, $UT_{(j+m)w} = h_1(st_{(j+m)w})$ *and* $CI_{(j+m)w} = h_2(k, st_{(j+m)w}) \oplus ind_m$, *where* $m = 1, 2, \dots, n$,

    - *Updates* ***List*** *to make* ***List****[w].c* $= (j+n)$ *and* ***List****[*$w_i$*].*$st_c = st_{(j+n)w}$

4. *The data owner sends the pair of* $EI_w = (UT_{(j+m)w}, CI_{(j+m)w})$ *to the investigator, where* $m = 1, 2, \dots, n$. *The pair of ciphertext of the index CI and its location in the server UT is represented as EI.*

5. *The cloud server stores* $CI_w$ *like this* $CD[UT_{(j+m)w}] = CI_{(j+m)wi}$, *where* $m = 1, 2, \dots, n$.

At this point the data owner has encrypted the *collected data* and the corresponding *keyword indexes* and uploaded all of it onto the cloud server. We now consider that the data has been compromised by a user. Consequently, the data owner has started reaching out to an investigator to conduct an investigation so that the perpetrator is identified. The process of inviting the investigator began by using the ***AuthToken*** algorithm. The steps taken by the data owner to give an authorization to the investigator are as follow.

***AuthToken($PK_r$, w)*** *takes as input the parameters ($PK_r$, w) public key of the investigator and the set of keywords, and performs the following:*

1. *The data owner selects a number $K \in Z_q{}^*$ randomly, and computes $AU_{1w} = K \cdot P$ and $Aw = K \cdot PK_r$.*
2. *The data owner computes $AU_{2w} = h_3(A_w, AU_{1w}) \oplus$ (**List**[w].c ‖ **List**[w].$st_c$).*
3. *The data owner sends the authorization token $A_w = (AU_{1w}, AU_{2w})$ to the investigator.*

Now the investigator is authorized to carry out a search. Subsequently, the investigator specifies a set of keywords that are believed to be helpful in obtaining the evidence required for the case. This is accomplished through the ***Trapdoor*** algorithm and the following steps are taken.

***Trapdoor($PK_s$, $SK_r$, $A_w$, w)*** *takes as input the parameters $PK_s = sP$, $SK_r = \alpha$ and the authorization token $A_w = (AU_{1w}, AU_{2w})$ of keyword w and performs the following:*

1. *The investigator computes $A_w{}^* = \alpha \cdot AU_{1w}$.*
2. *The investigator computes (**List***[w].c ‖ **List***[w].$st_c$) = $h_3(A_w, AU_{1w}) \oplus AU_{2w}$, and uses **List***[w] as the trapdoor of keyword w.*
3. *The investigator keeps the trapdoor $T_w = (c, st_{cw})$, where c = **List***[w].c, $st_{cw}$ = **List***[w].$st_c$ .*

The investigator is now ready with the trapdoor of the list of keywords to submit to the cloud server and to conduct the search. To accomplish the ***search*** process, the search algorithm is employed and the following steps were taken.

***Search(ED, $T_w$)*** *takes as input the ciphertext data ED, and the trapdoor $T_w = (c, st_{cw})$ of the keyword w sent by the investigator through a secure channel. The cloud server performs the following:*

1. *For each j, the cloud server computes $st_{(j-1)w} = f_{pk}(st_{jw})$, $UT_{jw} = h_1(st_{jw})$ and obtains the encrypted index $CI_{jw} = ED[UT_{jw}]$, where $j = c, c - 1, c - 2, \ldots, 1$*

2. *The cloud server returns $EI_w = \{CI_{1w}, CI_{2w}, \ldots, CI_{cw}\}$ to the investigator.*

Now the investigator has the result of the search, which is a set of encrypted data believed to be the evidence the investigator wanted for the identification of the user said to have breached the security policy of the data owner's company. With the private key ($SK_r$) the investigator can decipher the collected encrypted data returned by the server. As per the decision of the data-owner company's management, a disciplinary or a legal action can be taken against the user. This is based on the severity of the damage the user has caused the company.

Regarding the privacy issues that would have occurred with storing the proactively collected network data as plaintext, it would be prevented with the use of a searchable encryption that allows the investigator to work on certain keywords to retrieve evidence associated only with the suspected user. Likewise, the searchable encryption model proposed in this research protects the collected network from disgruntled employees who may be working for the cloud services to collude with perpetrators to delete or manipulate the evidence.

## 5.7 Chapter Summary

In this chapter, an encryption scheme that proactively protects collected network data from exposing private user information is proposed. The data owner may need to keep the investigation confidential from the cloud service provider who hosts the storage server of the data owner. Likewise, the data owner may not want to entrust the investigator with data that includes sensitive information. A searchable encryption scheme that allows the investigator to conduct a limited search using keywords approved by the data owner is proposed. The scheme was implemented in an experiment. The experiment proactively collected data from a network layer of an IoT ecosystem and stored it in a cloud. The feasibility of the scheme model was verified.

# Chapter 6: Conclusion and Future work

## 6.1 Conclusion

The main objective of the research was to find a solution for user privacy that can be breached during an investigation's analysis of data collected from IoT technologies. Several contributions have been made. First, we tried to unveil the issue of privacy that is normally associated with digital forensics that has been exacerbated by the increased use of IoT technologies. We have shown that existing standard digital-forensics models are not applicable to investigations involving IoT ecosystems. We suggested that there is a need for a new model that can contribute to the need for standardization of IoT-forensics tools and techniques. In explicating the issues and challenges associated with IoT forensics, we highlighted that the issue of user privacy is critically important and needs immediate attention.

Consequently, we aimed for the development of an IoT forensics model that preserves user privacy, which was based on design science research that was chosen as the methodology of this research. Next arose the need to identify a set of requirements that must be met by any model meant to be used in IoT forensics. As a result, a set of requirements that consists of 15 elements were identified. The research strategy adopted for the identification of the requirements was a document survey in which existing IoT-forensics-related literature was reviewed and the requirements derived. After the set of requirements was coined, a list of the requirements, with their corresponding descriptions, were submitting to a group of researchers for evaluation. The requirements were evaluated for usefulness, completeness, and effectiveness. The results were positive; all the experts rated the requirements as useful and complete, and rated them between effective and most effective. For further evaluation, the requirements were mapped to almost all the models primarily developed for IoT forensics. None of the models covered all the requirements. With this result, the need for a more sophisticated model was paramount.

Once the requirements were identified, we began the development of the new model. During the process of model development, a set of models selected from all domains of digital forensics were reviewed. The models were subjected to criteria for inclusion so that they could be used for the development of our model. Those that passed the criteria were then employed in the constitution of our model. Subsequently, activities proposed in each of the models were extracted and studied thoroughly in order to fit in the model. The relationship between the activities was studied and the model was designed, based on three design requirements. In the end, to remove subjectivity from the model-development process, the model, with some explicative knowledge, was submitted to experts for evaluation. The experts evaluated the model for its utility, usability, and ease of use. We used a semi-structured interview to obtain the results of the expert evaluations, which were generally positive. The experts suggested some activities to be included as adjustments, and these were accommodated in our model.

Based on the need for pre-investigative activities proposed in the model, it became important to propose a privacy-preserving mechanism that proactively protects collected potential digital evidence from any perpetrator but allows investigators to conduct a search in it. We proposed an encryption scheme that allows data owners to encrypt data stored in a cloud server and give permission to investigators to conduct searches with the acknowledgement of the owner. The proposed scheme uses a searchable encryption strategy that preserves private owner data. The owner has the capability of encrypting the data before it is sent to the cloud server and, when the need for an investigation arises, the data owner can give permission for the kind of search an investigators can perform. The cloud service provider does not have a say on the encrypted data. The provider cannot read the data and also cannot decipher it to manipulate the data.

## 6.2 Future Work

During our research we came across some important ideas that can be taken up as future work research.

1. Because the model was not tested against existing IoT forensics and cloud forensics tools, it would be useful to have used some internationally or legally supported tools; as this might produce a better understanding of which tools can be used in which part of the model.

2. The model can be subjected to further evaluation by first implementing it a real case and later submitting to USA experts to answer to the privacy protection rules familiar in USA.

3. The model provided a generic user privacy protection measures particularly those used in KSA. In this light, the model to cover EU privacy protection rules, it can be mapped to GDPR (General Data Protection Rules) and subsequently, areas where the model falls short of the GDPR and other privacy rules used in USA can be improved with conducting further research.

4. We did not instantiate any part of the model to show its usability as a leader of IoT-forensic tools; hence, in the future it will be valuable if some parts of the model are automated and implemented in an actual cloud-connected IoT environment.

5. The privacy-preserving searchable encryption scheme proposed in the research was only implemented as low fidelity prototype. But was not practically employed in a real case; developing the scheme and using it practically in a real case will be promising future work

## Reference

1. Assembly, U.G., *Universal declaration of human rights.* UN General Assembly, 1948.

2. Moussa, A.N., N. Ithnin, and A. Zainal, *CFaaS: bilaterally agreed evidence collection.* Journal of Cloud Computing, 2018. **7**(1): p. 1.

3. Hung, M., *Leading the iot, gartner insights on how to lead in a connected world.* Gartner Research, 2017: p. 1-29.

4. AlHogail, A., *Improving IoT Technology Adoption through Improving Consumer Trust.* Technologies, 2018. **6**(3): p. 64.

5. Stoyanova, M., et al., *A survey on the internet of things (IoT) forensics: challenges, approaches, and open issues.* IEEE Communications Surveys & Tutorials, 2020. **22**(2): p. 1191-1221.

6. Alabdulsalam, S., et al. *Internet of Things Forensics–Challenges and a Case Study*. in *IFIP International Conference on Digital Forensics*. 2018. Springer.

7. Mrdovic, S., *IoT Forensics*, in *Security of Ubiquitous Computing Systems*. 2021, Springer. p. 215-229.

8. Devi, M.G. and M.J. Nene, *Security Breach and Forensics in Intelligent Systems*, in *Information and Communication Technology for Intelligent Systems*. 2019, Springer. p. 349-360.

9. Atlam, H.F., et al., *Internet of things forensics: A review.* Internet of Things, 2020: p. 100220.

10. Nieto, A., R. Rios, and J. Lopez, *IoT-Forensics meets privacy: towards cooperative digital investigations.* Sensors, 2018. **18**(2): p. 492.

11. Dehghantanha, A. and K. Franke. *Privacy-respecting digital investigation*. in *2014 Twelfth Annual International Conference on Privacy, Security and Trust*. 2014. IEEE.

12. Pichan, A., M. Lazarescu, and S.T. Soh, *Towards a practical cloud forensics logging framework.* Journal of information security and applications, 2018. **42**: p. 18-28.

13. Perumal, S., N.M. Norwawi, and V. Raman. *Internet of Things (IoT) digital forensic investigation model: Top-down forensic approach methodology*. in *2015 Fifth International Conference on Digital Information Processing and Communications (ICDIPC)*. 2015. IEEE.

14. Babun, L., et al., *IoTDots: A Digital Forensics Framework for Smart Environments.* arXiv preprint arXiv:1809.00745, 2018.

15. Kebande, V.R. and I. Ray. *A generic digital forensic investigation framework for internet of things (iot)*. in *2016 IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud)*. 2016. IEEE.

16. Yaqoob, I., et al., *Internet of things forensics: Recent advances, taxonomy, requirements, and open challenges.* Future Generation Computer Systems, 2019. **92**: p. 265-275.

17. Nieto, A., R. Roman, and J. Lopez*, Digital witness: Safeguarding digital evidence by using secure architectures in personal devices.* IEEE Network, 2016. **30**(6): p. 34-41.

18. Nieto, A., R. Rios, and J. Lopez. *A methodology for privacy-aware IoT-forensics*. in *2017 IEEE Trustcom/BigDataSE/ICESS*. 2017. IEEE.

19. Nieto, A., R. Rios, and J. Lopez. *Digital witness and privacy in IoT: Anonymous witnessing approach*. in *2017 IEEE Trustcom/BigDataSE/ICESS*. 2017. IEEE.

20. Le, D.-P., et al. *BIFF: a blockchain-based IoT forensics framework with identity privacy*. in *TENCON 2018-2018 IEEE Region 10 Conference*. 2018. IEEE.

21. Uddin, M.A., et al., *A Survey on the Adoption of Blockchain in IoT: Challenges and Solutions.* Blockchain: Research and Applications, 2021: p. 100006.

22. Zulkipli, N.H.N. and G.B. Wills, *An Exploratory Study on Readiness Framework in IoT Forensics.* Procedia Computer Science, 2021. **179**: p. 966-973.

23. Ariffin, K.A.Z. and F.H. Ahmad, *Indicators for Maturity and Readiness for Digital Forensic Investigation in Era of Industrial Revolution 4.0.* Computers & Security, 2021: p. 102237.

24. Dresch, A., D.P. Lacerda, and J.A.V. Antunes, *Proposal for the conduct of design science research*, in *Design Science Research*. 2015, Springer. p. 117-127.

25. Alismail, S., H. Zhang, and S. Chatterjee. *A Framework for Identifying Design Science Research Objectives for Building and Evaluating IT Artifacts*. in *International Conference on Design Science Research in Information System and Technology*. 2017. Springer.

26. Venable, J., J. Pries-Heje, and R. Baskerville. *A comprehensive framework for evaluation in design science research*. in *International Conference on Design Science Research in Information Systems*. 2012. Springer.

27. Hevner, A. and S. Chatterjee, *Design science research in information systems*, in *Design research in information systems*. 2010, Springer. p. 9-22.

28. Peffers, K., et al., *A design science research methodology for information systems research.* Journal of management information systems, 2007. **24**(3): p. 45-77.

29. Halboob, W., et al., *Privacy levels for computer forensics: toward a more efficient privacy-preserving investigation.* Procedia Computer Science, 2015. **56**: p. 370-375.

30. Aminnezhad, A., A. Dehghantanha, and M.T. Abdullah, *A survey on privacy issues in digital forensics.* International Journal of Cyber-Security and Digital Forensics, 2012. **1**(4): p. 311-324.

31. Law, F.Y., et al. *Protecting digital data privacy in computer forensic examination*. in *2011 Sixth IEEE International Workshop on Systematic Approaches to Digital Forensic Engineering*. 2011. IEEE.

32. Hou, S., et al. *Privacy preserving confidential forensic investigation for shared or remote servers*. in *2011 Seventh International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. 2011. IEEE.

33. Reddy, K. and H. Venter. *A forensic framework for handling information privacy incidents*. in *IFIP International Conference on Digital Forensics*. 2009. Springer.

34. Ajijola, A., P. Zavarsky, and R. Ruhl. *A review and comparative evaluation of forensics guidelines of NIST SP 800-101 Rev. 1: 2014 and ISO/IEC 27037: 2012*. in *World Congress on Internet Security (WorldCIS-2014)*. 2014. IEEE.

35. Caron, X., et al., *The Internet of Things (IoT) and its impact on individual privacy: An Australian perspective.* Computer Law & Security Review, 2016. **32**(1): p. 4-15.

36. Malhotra, N.K., S.S. Kim, and J. Agarwal, *Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model.* Information systems research, 2004. **15**(4): p. 336-355.

37. Thierer, A., *The pursuit of privacy in a world where information control is failing.* Harv. JL & Pub. Pol'y, 2013. **36**: p. 409.

38. Moussa, A.N., N.B. Ithnin, and O.A. Miaikil. *Conceptual forensic readiness framework for infrastructure as a service consumers*. in *2014 IEEE Conference on Systems, Process and Control (ICSPC 2014)*. 2014. IEEE.

39. Curzon, J., A. Almehmadi, and K. El-Khatib, *A survey of privacy enhancing technologies for smart cities.* Pervasive and Mobile Computing, 2019.

40. Fang, W., et al., *A survey of big data security and privacy preserving.* IETE Technical Review, 2017. **34**(5): p. 544-560.

41. Raghunathan, B., *The complete book of data anonymization: from planning to implementation*. 2013: Auerbach Publications.

42. Domingo-Ferrer, J., D. Sánchez, and J. Soria-Comas, *Database anonymization: privacy models, data utility, and microaggregation-based inter-model connections.* Synthesis Lectures on Information Security, Privacy, & Trust, 2016. **8**(1): p. 1-136.

43. Li, H., K. Muralidhar, and R. Sarathy, *The Effectiveness of Data Shuffling for Privacy-Preserving Data Mining Applications.* Journal of Information Privacy and Security, 2012. **8**(2): p. 3-17.

44. Abawajy, J.H., M.I.H. Ninggal, and T. Herawan, *Privacy preserving social network data publication.* IEEE communications surveys & tutorials, 2016. **18**(3): p. 1974-1997.

45. Aminnezhad, A. and A. Dehghantanha, *A survey on privacy issues in digital forensics.* International Journal of Cyber-Security and Digital Forensics (IJCSDF), 2014. **3**(4): p. 183-199.

46. Cebe, M., et al., *Block4forensic: An integrated lightweight blockchain framework for forensics applications of connected vehicles.* IEEE Communications Magazine, 2018. **56**(10): p. 50-57.

47. Ryu, J.H., et al., *A blockchain-based decentralized efficient investigation framework for IoT digital forensics.* The Journal of Supercomputing, 2019: p. 1-16.

48. Palmer, G. *A road map for digital forensic research*. in *First Digital Forensic Research Workshop, Utica, New York*. 2001.

49. Van Staden, F. and H.S. Venter. *Adding digital forensic readiness to electronic communication using a security monitoring tool*. in *Information Security South Africa (ISSA), 2011*. 2011. IEEE.

50. Reith, M., C. Carr, and G. Gunsch, *An examination of digital forensic models.* International Journal of Digital Evidence, 2002. **1**(3): p. 1-12.

51. Koroniotis, N., N. Moustafa, and E. Sitnikova, *Forensics and Deep Learning Mechanisms for Botnets in Internet of Things: A Survey of Challenges and Solutions.* IEEE Access, 2019. **7**: p. 61764-61785.

52. Henry, P., J. Williams, and B. Wright, *The SANS Survey of Digital Forensics and Incident Response.* A white paper, SANS analyst program, 2013.

53. Carrier, B. and E.H. Spafford. *An event-based digital forensic investigation framework*. in *Digital forensic research workshop*. 2004.

54. Ieong, R.S., *FORZA–Digital forensics investigation framework that incorporate legal issues.* digital investigation, 2006. **3**: p. 29-36.

55. Kohn, M.D., M.M. Eloff, and J.H. Eloff, *Integrated digital forensic process model.* Computers & Security, 2013. **38**: p. 103-115.

56. Wilkinson, S. and D. Haagman, *Good practice guide for computer-based electronic evidence.* Association of Chief Police Officers, 2010.

57. Conti, M., et al., *Internet of Things security and forensics: Challenges and opportunities*. 2018, Elsevier.

58. Herzberg, B., D. Bekerman, and I. Zeifman, *Breaking down mirai: An IoT DDoS botnet analysis.* Incapsula Blog, Bots and DDoS, Security, 2016.

59. Bharadwaj, N.K. and U. Singh, *Acquisition and Analysis of Forensic Artifacts from Raspberry Pi an Internet of Things Prototype Platform*, in *Recent Findings in Intelligent Computing Techniques*. 2019, Springer. p. 311-322.

60. Liu, Y.-N., et al., *Privacy-preserving raw data collection without a trusted authority for IoT.* Computer Networks, 2019. **148**: p. 340-348.

61. Veber, J. and Z. Smutny. *Standard ISO 27037: 2012 and collection of digital evidence: Experience in the Czech Republic*. in *European Conference on Cyber Warfare and Security*. 2015. Academic Conferences International Limited.

62. Li, S., et al., *IoT forensics: Amazon echo as a use case.* IEEE Internet of Things Journal, 2019.

63. Croft, N.J. and M.S. Olivier, *Sequenced release of privacy-accurate information in a forensic investigation.* Digital Investigation, 2010. **7**(1-2): p. 95-101.

64. Janarthanan, T., M. Bagheri, and S. Zargari, *IoT Forensics: An Overview of the Current Issues and Challenges.* Digital Forensic Investigation of Internet of Things (IoT) Devices, 2021: p. 223-254.

65. Oriwoh, E., et al. *Internet of things forensics: Challenges and approaches*. in *9th IEEE International Conference on Collaborative computing: networking, Applications and Worksharing*. 2013. IEEE.

66. CHEN, F., et al., *IoT Cloud Security Review: A Case Study Approach Using Emerging Consumer-Oriented Applications.* 2021.

67. Liu, J., R. Sasaki, and T. Uehara. *Towards a Holistic Approach to Medical IoT Forensics*. in *2020 IEEE 20th International Conference on Software Quality, Reliability and Security Companion (QRS-C)*. 2020. IEEE.

68. Alhassan, J., et al. *Comparative evaluation of mobile forensic tools*. in *International Conference on Information Theoretic Security*. 2018. Springer.

69. Barmpatsalou, K., et al., *Mobile Forensic Data Analysis: Suspicious Pattern Detection in Mobile Evidence.* IEEE Access, 2018. **6**: p. 59705-59727.

70. Kebande, V.R., et al., *Digital forensic readiness intelligence crime repository.* Security and Privacy, 2021: p. e151.

71. Adams, R., *The advanced data acquisition model (ADAM): a process model for digital forensic practice*. 2012, Murdoch University.

72. Ayers, R.P., S. Brothers, and W. Jansen, *Guidelines on mobile device forensics*. 2014.

73. Ren, W., et al., *Privacy-preserving using homomorphic encryption in Mobile IoT systems.* Computer Communications, 2021. **165**: p. 105-111.

74. Alexakos, C., et al., *Enabling Digital Forensics Readiness for Internet of Vehicles.* Transportation Research Procedia, 2021. **52**: p. 339-346.

75. Sandvik, J.-P., K. Franke, and A. Årnes, *Towards a Generic Approach of Quantifying Evidence Volatility in Resource Constrained Devices*, in *Digital Forensic Investigation of Internet of Things (IoT) Devices*. 2021, Springer. p. 21-45.

# Appendix A - Evaluation Form

## EVALUATION OF PRIVACY-PRESERVING IoT-FORENSICS MODEL IN TERMS OF PRACTITIONER PERSPECTIVES

**Student Name:**

**Institution:**

**Faculty:**

**Department:**

**Supervisor:**

| | |
|---|---|
| **Evaluator's Name** | |
| **Position** | |
| **Organization** | |
| **Work Experience** | |
| **Signature** | |

## Overview

The representation of the model with ~~the~~ three layers is in BPMN. BPMN process diagrams are incorporated within the model to define the flow of the steps of an investigation in each of the layers. Subsequently, the model was developed by employing BPMN as it supports instantiation of the model to tailor for different IoT-ecosystem scenarios, thereby making the model generic. For the model to assist investigators ~~to~~ with applying the concepts and processes given in the model, investigators should and must do each layer in conjunction with other layers at a level of detail that permits admissibility of a case.

Each layer of the model is represented in BPMN and discussed separately. The common processes that transverse between layers include abiding to legal terms, evidence preservation, and chain of custody. This is to ensure that all activities associated with the collection of evidence and subsequently the transporting and storing of the data are protected, as there is the potential ~~of~~ that the whole process of investigation could be ordered to be repeated by court. The model does not require investigators to be specialists in performing IoT forensics; rather an investigator can easily follow the explicit steps provided in the model. Nevertheless, an investigator must observe the requirements so that their actions are auditable through clear documentation, repeatable where possible using the same tools on the same IoT devices under the same conditions, the same result reproducible with the use of different tools on the same IoT devices, and justifiable.

The premises where the IoT devices are located is where investigations usually start. The steps proposed in this layer include preparation, evidence source identification and preservation, collection and acquisition, examination and analysis, and reporting and presentation. The privacy protection protocols employed in each of the steps are discussed accordingly. However, before any privacy measures were added to the steps, the main concepts of the steps were briefly discussed.

Please read the requirements and the discussions given for the model and then answer the questions below. The cloud-forensic-process model developed in this research is shown in the figures below. Your name is purely for my own administration and will not be disclosed to anyone. If you would like other aspects to not be disclosed, please let me know.

## Model Requirements

Please consider that these requirements are not meant as a minimum baseline to be achieved, but rather aim for preparing a privacy-preserving IoT-forensic investigation as much as possible. In your view, some may add only a very limited prospect, but still add something to the level of user-privacy protection nonetheless.

RQ1: Digital Forensic Readiness (DFR) processes include: extraction of digital evidence, parsing forensic logs, digital preservation, creation of hash values, evidence storage, log analysis and characterization, and a readiness report.

RQ2: The volume of IoT data captured by sensors and smart devices from networks and the cloud complicates the identification of relevant data.

RQ3: The parties should be held responsible for their actions by providing proof of integrity.

RQ4: The system should have minimum overhead on endpoints since it includes multiple parties that may have different capabilities and resources.

RQ5: Consent and choice where an IoT consumer should give consent on the collection of his data.

RQ6: The purpose legitimacy and specification requirement promotes the idea that the consumer has to be informed about the reason for the data collection.

RQ7: The collection limitation requirement refers to the collection of data strictly relevant for the case at hand.

RQ8: The data minimization requirement answers to the large volume of data that can be collected from the IoT ecosystem and hence promotes the reduction of the data to its minimum volume possible.

RQ9: The use, retention, and disclosure limitation requirement are to ensure that data collected must not be used for a purpose other than the one originally specified.

RQ10: The accuracy and quality requirement~~s~~ asks that an investigator follow the data collection process in a trusted and admissible way.

RQ11: The openness, transparency, and notification requirement reinforces that a consumer must be informed of the procedure, policies, and practices of the forensic analysis that his data is going to be subjected to.

RQ12: The individual participation and access requirement proposed by the researchers require that a consumer must have access to his data throughout the process of investigation.

RQ13: The accountability requirement promotes that the investigator has to follow the privacy policies set forth for the collection and analysis of the evidence.

RQ14: The information security controls requirement protects collected personal data from unauthorized access, loss, and modification.

RQ15: The compliance requirement incorporates the implementation of an auditing mechanism to ensure that the whole process of investigation complies with the investigative and privacy principles.

## Proposed Cloud Forensic Process Model

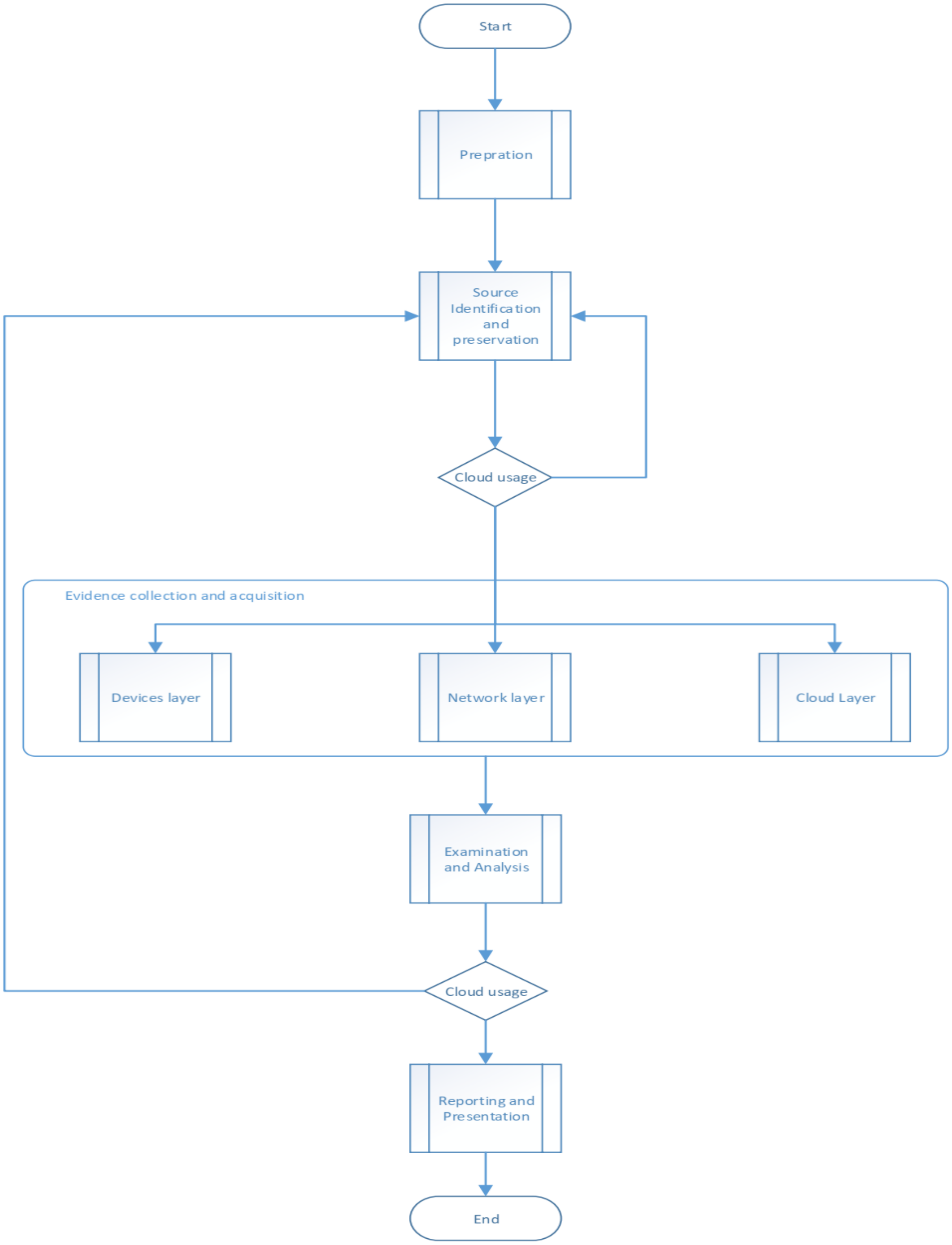

**Figure 1**. High-level of the Model

The bigger picture or the higher-level view of the model is illustrated in Figure 1. The processes were appropriately designed by cross-referencing with the ACPO principles. This new IoT-forensics model uses the readiness, live, and postmortem components together[75]. Figures 2–4 explicitly show processes that support~~s~~ the workflow by establishing the activities and tasks to ensure that some of the highly required activities are not bypassed, switched, or not followed.

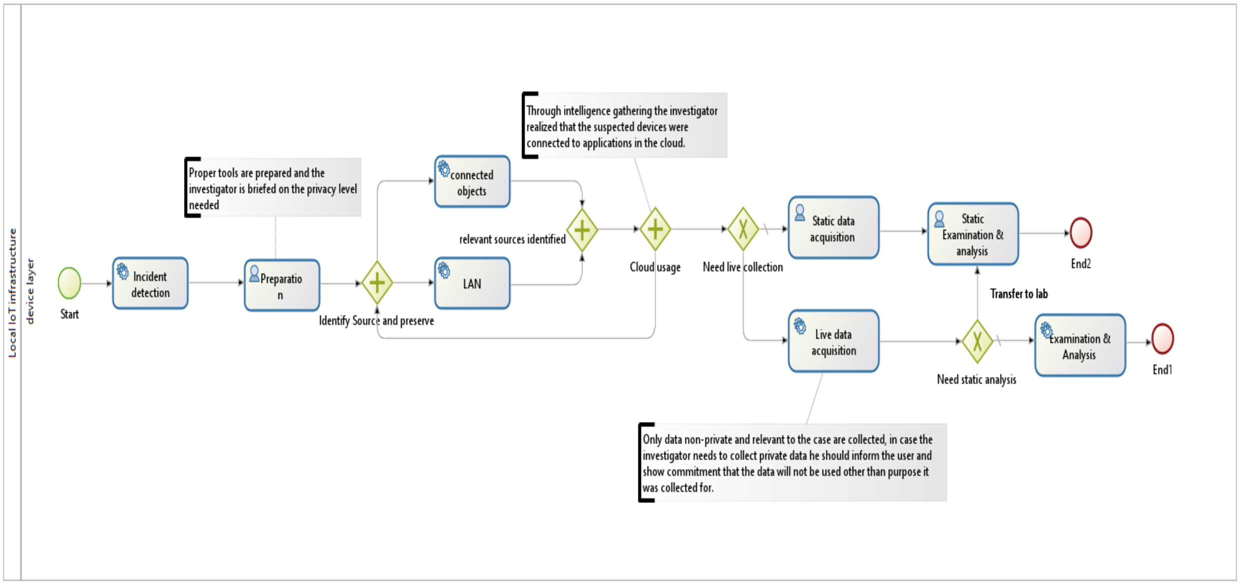
Local IoT infrastructure
device layer
Start
Incident detection
Preparatio n
Proper tools are prepared and the investigator is briefed on the privacy level needed
Identify Source and preserve
connected objects
LAN
relevant sources identified
Through intelligence gathering the investigator realized that the suspected devices were connected to applications in the cloud.
Cloud usage
Need live collection
Static data acquisition
Static Examination & analysis
End2
Transfer to lab
Live data acquisition
Need static analysis
Examination & Analysis
End1
Only data non-private and relevant to the case are collected, in case the investigator needs to collect private data he should inform the user and show commitment that the data will not be used other than purpose it was collected for.

**Figure 2:** Device layer forensics

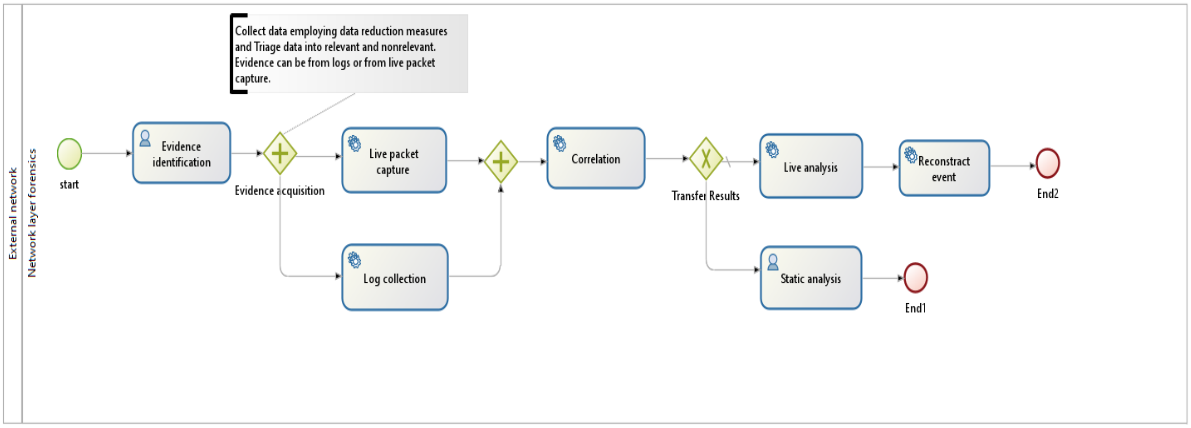

**Figure 3:** Network layer forensics

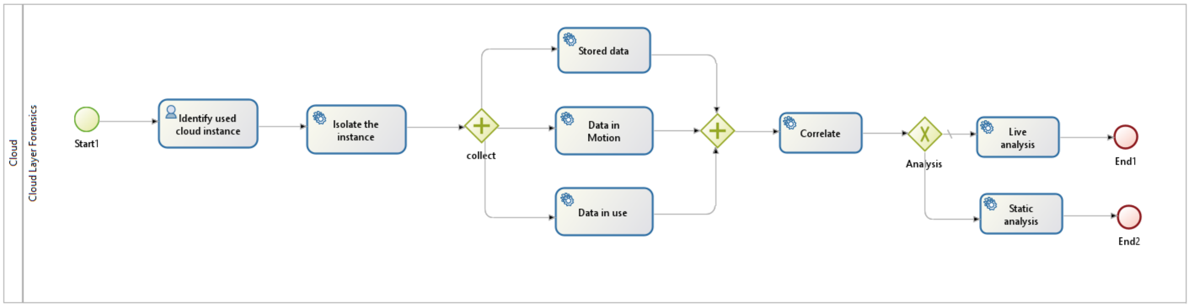

**Figure 4:** Cloud layer forensic

The documentation and preservation process are included in the model as a continuous process. This process includes the recording of documentation, preservation of the chain of custody, and preservation of digital evidence made as accurately as possible throughout the entire investigation. Likewise, the legal and contractual review process is another continuous process that refers to the underlying contractual agreements and the legal jurisdictions the cloud consumers and providers are subject to and supposed to adhere to in order to withstand legal scrutiny.

Regarding **Fig.2**, IoT devices are prepared to acquire only security incident-related data instead of everything that may include personal user information that may cause a breach of privacy when investigators are gathering evidence. The device layer in fact contains all physical resources that collect and control data. Thus~~,~~ the need~~s~~ for unique approaches ~~of~~ to protect the privacy of individuals and organizations.

Regarding **Fig.3,** there is a need to meet the connectivity demand of IoT technology because network data has increased in terms of scale and complexity due to the increased adoption of IoT technologies. IoT devices send a huge amount of data regardless of the user's will. Hence, mapping between traffic received at some point in the network and an IoT device of origin is challenging. This is the main issue that triggers a privacy concern when an investigator seizes and investigates devices of innocent users, thereby exposing their private or confidential information. The privacy-protection preparation needed in this network layer may therefore include providing a mechanism of attribution where network data can be linked to IoT devices and their users.

Regarding **Fig.4,** due to the huge volume of data and the distributed nature of the cloud, it is difficult to collect evidence and associate it with a particular IoT device. For example, several devices may use the same cloud instance and therefore collecting data from such an instance may cause a breach to the private data of some innocent IoT-device users. Hence, the privacy-protection preparation needed in this cloud layer may include maintaining proper communication channels with the cloud service provider so that investigators may ask the provider to conduct a dedicated search and collect relevant information.

## Questions ~~&~~ and Answers

Q1: Are the elements defined in the requirements useful for privacy-preserving IoT-forensic models?

| | | |
|---|---|---|
| Answering | Yes □ | No □ |
| Comments | | |

Q2: Are there any missing elements in the defined requirements that you think would be important for IoT forensics? If yes, please mention them.

| | | |
|---|---|---|
| Answering | Yes □ | No □ |
| Comments | | |

Q3: In the table below, please indicate on a scale from 1-5 how effective you think each requirement is with regard~~s~~ to user privacy-preserving IoT-forensic analysis. *(1 being least effective and 5 being most effective)*

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Requirements | RQ1 | RQ2 | RQ3 | RQ4 | RQ5 | RQ6 | RQ7 | RQ8 | RQ9 | RQ10 | RQ11 | RQ12 | RQ13 | RQ14 | RQ15 |
| Effectiveness | | | | | | | | | | | | | | | |

Q4: In reviewing this model, please identify any aspect that would not be representative in the process as carried out in your environment when investigating IoT environments. (Utility)

| Remove | Reason for removal |
|---|---|
| | |

Q5: In reviewing this model, please identify any aspects of your IoT-investigation activities that are not covered. (Utility)

| Add | Reason for Adding |
|---|---|
| | |

Q6: Please identify any aspects of the model that you feel could be improved. (Utility)

| Improve | Reason for Improving |
|---|---|
| | |

Q7: Please rate this model in terms of the usability of the process model from 1 to 10 (*with 1 being very difficult to use and adopt and 10 being very easy to use and adopt*). (Usability)

| Rate | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Usability | □ | □ | □ | □ | □ | □ | □ | □ | □ | □ |

If you have any further comments or suggestions about this model and its components, please feel free to provide them below.

| |
|---|
| |

# Appendix B - Python Source Code

## Python Build Index code lines

```
import pandas as pd
from Crypto.PublicKey import RSA
from Crypto.Hash import MD5
from Crypto.Random import random
import numpy as np
import time


def build_trapdoor(MK, keyword):
    keyword_index = MD5.new()
    keyword_index.update(str(keyword))
    ECB_cipher = RSA.new(MK, RSA.MODE_ECB)
    return ECB_cipher.encrypt(keyword_index.digest())


def build_codeword(ID, trapdoor):
    ID_index = MD5.new()
    ID_index.update(str(ID))
    ECB_cipher = RSA.new(trapdoor, RSA.MODE_ECB)
    return ECB_cipher.encrypt(ID_index.digest()).encode("hex")


def build_index(MK, ID, keyword_list):
    secure_index = [0] * len(keyword_list)
    for i in range(len(keyword_list)):
        codeword = build_codeword(ID, build_trapdoor(MK, keyword_list[i]))
        secure_index[i] = codeword
    random.shuffle(secure_index)
    return secure_index

def searchable_encryption(raw_data_file_name, master_key,
keyword_type_list):
    raw_data = pd.read_csv(raw_data_file_name)
    features = list(raw_data)
    raw_data = raw_data.values

    keyword_number = [i for i in range(0, len(features)) if features[i] in
keyword_type_list]

    index_header = []
    for i in range(1, len(keyword_type_list) + 1):
        index_header.append("index_" + str(i))

    document_index = []
    start_time = time.time()
    for row in range(raw_data.shape[0]):
        record = raw_data[row]
        record_keyword_list = [record[i] for i in keyword_number]
        record_index = build_index(master_key, row, record_keyword_list)
        document_index.append(record_index)

    time_cost = time.time() - start_time
    print time_cost
    document_index_dataframe = pd.DataFrame(np.array(document_index), 
columns=index_header)
    document_index_dataframe.to_csv(raw_data_file_name.split(".")[0] + 
"_index.csv")


if __name__ == "__main__":

    document_name = raw_input("Please input the file to be encrypted:  ")

    Public_key_file_name = raw_input("Please input the file stored the 
public key:  ")
    Public_key = open(Public_key_file_name).read()
    if len(Public_key) > 16:
        print "the length of public key is larger than 16 bytes, only the 
first 16 bytes are used"
        Public_key = bytes(public_key[:16])

    keyword_list_file_name = raw_input("please input the file stores 
keyword type:  ")
    keyword_type_list = open(keyword_list_file_name).read().split(",")

    searchable_encryption(document_name, public_key, keyword_type_list)

    print "Finished"
```

## Python Trapdoor Code lines

```
from Crypto.Publickey import RSA
from Crypto.Hash import MD5


def build_trapdoor(MK, keyword):
    keyword_index = MD5.new()
    keyword_index.update(str(keyword))
    ECB_cipher = RSA.new(MK, RSA.MODE_ECB)
    return ECB_cipher.encrypt(keyword_index.digest())

if __name__ == "__main__":

    keyword = raw_input("Please input the keyword you want to search:  ")

    public_key_file_name = raw_input("Please input the file stored the 
pubilc key:  ")
    public_key = open(public_key_file_name).read()
    if len(public_key) > 16:
        print "the length of master key is larger than 16 bytes, only the 
first 16 bytes are used"
        public_key = bytes(public_key[:16])


    trapdoor_file = open(keyword + "_trapdoor", "w+")
    trapdoor_of_keyword = build_trapdoor(public_key, keyword)
    trapdoor_file.write(trapdoor_of_keyword)
    trapdoor_file.close()
```

## Python search code line

```
import pandas as pd
from Crypto.Publickey import RSA
from Crypto.Hash import MD5
# import time

def build_codeword(ID, trapdoor):
    ID_index = MD5.new()
    ID_index.update(str(ID))
    ECB_cipher = RSA.new(trapdoor, RSA.MODE_ECB)
    return ECB_cipher.encrypt(ID_index.digest()).encode("hex")

def search_index(document, trapdoor):
    search_result = []
    data_index = pd.read_csv(document)
    data_index = data_index.values
    # start_time = time.time()
    for row in range(data_index.shape[0]):
        if build_codeword(row, trapdoor) in data_index[row]:
            search_result.append(row)

    # print time.time() - start_time
    return search_result

if __name__ == "__main__":

    index_file_name = raw_input("Please input the index file you want to
search:  ")
    keyword_trapdoor = raw_input("Please input the file stored the trapdoor
you want to search:  ")
    keyword_trapdoor = open(keyword_trapdoor).read().strip()
    search_result = search_index(index_file_name, keyword_trapdoor)
    print "The identifiers of files that contain the keyword are: \n",
search_result
```