# Digital Evidence Challenges in the Internet of Things

R.C.Hegarty[1], D.J.Lamb[2] and A.Attwood[3]

[1]School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University, John Dalton Building, Chester Street, Manchester, UK
[2]School of Computing & Mathematical Sciences, [3]School of Engineering, Technology and Maritime Operations, Liverpool John Moores University James Parsons Building, Byrom Street, Liverpool, UK
e-mail: R.Hegarty@mmu.ac.uk; {D.J.Lamb;A.J.Attwood}@ljmu.ac.uk

## Abstract

The implementation of the Internet of Things will result in the connection of tens of billions of wireless devices to the Internet. These devices will form an intelligent substrate pervading all aspects of life. From intelligent home control to advanced city management systems, devices will sense their environment as well as interconnect and communicate with each other to form intelligent smart spaces. Individually and collectively, these devices produce and consume large amounts of personally sensitive data. This new environment provides a rich set of data sources; when used in conjunction with one another, they can greatly inform a historical situation that may have occurred with little or no reliable human witness evidence. However, this deeply pervasive environment will provide challenges to the various agencies that will need to interact with this new technology. This paper establishes the fundamental overarching challenges the IoT poses to digital forensics, and identifies the key areas that solutions should target.

## Keywords

Internet of Things, Digital Investigations, Cloud Computing, Digital Forensics

## 1. Introduction: The Internet of Things

The Internet of Things (IoT), in the context of this paper, describes a world where many otherwise ordinary devices are uniquely identifiable, addressable and contactable via the Internet. Such *things* may be little more than a sensor or actuator augmented with basic transceiver electronics to manage connectivity requirements. Alternatively, these *things* may be sophisticated and comparable to typical consumer electronics devices, providing local computational functionality – such as a set-top Digital Video Recorder, or the headline-grabbing intelligent fridges commandeered in the infamous "Spam Fridge" attacks (Chirgwin, 2014).

As such, some *things* may have a degree of server functionality and respond readily to incoming requests and queries, whereas less sophisticated devices may simply generate and transmit their output data on certain triggers. They may take their power from a mains, self-generating or sustainable supply, or may be tightly constrained low-power battery operated devices with a limited power lifespan.

Additionally, their connectivity state – and therefore participation in the IoT – may be temporary or sporadic; based on the availability or status of power supply, the presence of a compatible *thing*-to-Internet gateway, or only triggered by *thing*-specific stimuli. The devices may therefore spend much of their time in a very low power state; idle or disconnected.

However, regardless of their appearance and power source, their placement – and looking to the future, their ubiquity – will allow them to collect and potentially store and process tremendous amounts of data. This data may not in itself be directly or deliberately personally identifying. Its combination and correlation with other datasets, both virtual and real, can provide significant insight into personal activity and environmental conditions.

This paper identifies the challenges IoT poses to established digital forensic procedures, and the areas in which solutions should be targeted. Section 2 provides an overview of the general digital forensics literature to illustrate where the challenges posed by the IoT interact with existing digital forensics models. In addition to this, recent works on the challenges posed by the IoT to digital forensics are reviewed to establish the state of the art research in this area. Section 3 describes the challenges posed by the IoT and some approaches to overcoming them. Section 4 identifies the key issues and emerging requirements of IoT investigations. Section 5 describes the future research directions and work required to develop solutions to the challenges posed to IoT investigations.

## 2. Related Work

### 2.1. Digital Forensics

Law enforcement agencies, private organisations and even individuals are familiar with the combination of digital and physical evidence resulting from forensic investigations. As electronic devices become pervasive this trend looks set to continue. For example it is commonplace for the police to check with mobile network operators, whether drivers were using a mobile telephone, when investigating road traffic incidents (GM Casualty Reduction Partnership, 2014) (Haines, 2007). Internet of Things devices augment today's digital data environment with potentially significant personal and context-setting data feeds. We identify the requirements for investigations in the IoT by first considering the seminal work of McKemmish (McKemmish, 1999) that describes Digital Forensics as:

*"The process of identifying, preserving, analysing and presenting digital evidence in a manner that is legally acceptable"*

The application of this definition to investigation in the IoT raises a number of challenges. To develop an understanding of these challenges and their place in the digital forensic process we consider the two predominant models for digital forensic investigations. Table  provides an overview of the models proposed by McKemmish (McKemmish, 1999) and NIST (Kent, Chevalier, Grance, Dang, & Kent, 2006).

| Stage | McKemmish | Kent |
|---|---|---|
| 1 | **Identification**, in this stage the location and format of evidence is identified to enable an appropriate mechanism to be determined for the purpose of recovering evidence. Digital evidence can be found in a myriad of places; computers, mobile phones, smart cards, set top boxes etc. | **Collection**, encompasses identification, preservation and acquisition of relevant evidence |
| 2 | **Preservation**, it is imperative that evidence is preserved as in many cases it will be the subject of judicial scrutiny. In some circumstances changes to data are unavoidable. In these cases change should be minimised and the process causing the change documented along with an explanation/justification of why the change was required. | **Examination**, uses automated and manual tools to extract data of interest. |
| 3 | **Analysis,** consists of the extraction, processing and interpretation of digital evidence. It forms the main element of forensic computing. Following extraction, processing is often required to make data human readable. Processing of extracted data may be part of the extraction stage or a separate stage in its own right. | **Analysis**, the derivation of useful information from the results of the examination stage. |
| 4 | **Presentation**, the final stage of the process involves a presentation of both the evidence and the process by which the evidence was gathered along with the presenter's qualifications. | **Reporting**, is concerned with the preparation and presentation of the evidence and forensic analysis process. |

**Table 1: Digital Forensics Process Models**

In the following section, we survey the literature relating to IoT forensics. We then revisit the investigatory stages detailed in Table , to consider the challenges specific to each stage of the investigatory process.

### 2.2. IoT Forensics

The identification and preservation of evidence in digital forensic investigations in emerging environments has always presented a challenge. Taylor et al (Taylor, Haggerty, Gresty, & Hegarty, 2010) suggested that the standards by which evidence is judged in digital forensic investigations may have to be altered to accommodate the changing nature of digital evidence from a cloud computing environment. We believe the same proposition holds true for IoT investigations.

Oriwoh et al (Oriwoh, Jazani, Epiphaniou, & Sant, 2013) identified preservation as a challenge, and suggested that devices undergoing investigation should not be turned off to preserve the modified, created and accessed times of files. Their assertion is likely drawn from conventional digital forensic investigations; however, the situation is much more complex in IoT investigations. Thought must be given to the limited

resources available on devices, leaving the devices running at the scene of an incident will use power, and more importantly may result in overwriting of stored data due to constrained storage capabilities. Therefore, consideration is required to determine whether devices should be powered off or left running. We study this challenge in further detail in section 3.

The extraction and preservation of data from devices and services running in the IoT will present challenges. Proprietary data formats, protocols, and physical interfaces all complicate the process of evidence extraction (Miorandi, Sicari, De Pellegrini, & Chlamtac, 2012). Some schemes distribute information to adjacent nodes within the same topology or to external cloud services. In these scenarios, investigators need to be able to identify the benefit to the investigation in extracting data from other nodes, base stations, or cloud services (Attwood, Merabti, & Abuelmaatti, 2011). This approach could be viable and may overcome some of the challenges associated with extracting data from devices with limited storage. The data stored and processed in the IoT can be of a sensitive nature. We posit that the way in which large corporations aggregate and process data, may be the subject of future digital forensic investigations.

Oriwoh et al (Oriwoh et al., 2013) identify the challenge posed by devices crossing the boundaries of jurisdictions; while we agree this is a challenge. It is highly likely that data in transit between IoT devices and globally distributed cloud computing platforms cross these boundaries on a far more frequent basis. In Section 3 we question whether the emphasis of investigations should be on devices or data, and whether devices may be viewed as a metadata aspect of the data or if the reverse is true.

## 3. IoT Forensic Challenges & Approaches

The IoT will undoubtedly provide a richer source of evidence from the physical world than conventional computer systems. The way in which IoT is realising Zelkha et al's vision of ambient intelligence (Zelkha, Epstein, Birrell, & Dodswoth, 1998) means that environments are beginning to react to the user's requirements, without the need for conscious interaction by the user. As a result, IoT environments are likely to contain contextual evidence of which the perpetrators are simply oblivious. This paradigm shift means that digital investigations will increasingly encounter evidence from events taking place in the physical world.

The four main phases of digital forensics investigation from Table 1 face a number of challenges from the IoT. We discuss the implications of the IoT for each phase under the headings below and identify areas in which solutions should be targeted

### 3.1. Identification

Detecting the presence of IoT systems poses challenges to digital forensic investigations, as does the identification of a particular user's data. This raises the question of how to carry out what law enforcement term "search & seizure" when it is not apparent where the data being investigated is being stored, or where the data came from.

A potential solution to identification of data may be the integration of IoT device data into Building Information Modelling ("National BIM Standards," 2013);

*Building Information Modelling (BIM) is a digital representation of physical and functional characteristics of a facility. A BIM is a shared knowledge resource for information about a facility forming a reliable basis for decisions during its life-cycle; defined as existing from earliest conception to demolition.*

By combining the information about the IoT capabilities of a building or structure, it may be possible to answer the questions of; where has the information come from? Where is the information stored? It is also crucial to identify in what format the data is stored or encoded. This would narrow the scope of the investigation, and enable the selection of features or data that identifies an individual user from a much smaller data set. A composite picture of the data gathered about an individual user could be constructed from the data stored or forwarded by the buildings they have inhabited.

### 3.2. Preservation

There are established procedures in place to capture volatile evidence before it becomes unavailable, for example first responders can create memory dumps prior to a machine being shut down (Thomas, Sherly, & Dija, 2013). Evidence volatility in the IoT is much more complex; data may be stored locally by a *thing*, in which case the lifespan of the data before it is overwritten or compressed using a lossy technique is finite. The data from a *thing* may be transferred and consumed by another thing or a local ad-hoc network of *things*, alternatively it may transferred to the cloud for aggregation and processing.

The transfer and aggregation of data/evidence presents a challenge when securing the chain of evidence. In order to overcome this challenge and leverage the resilient nature of data in IoT in digital investigations, techniques are required to track and filter the transit of data across an IoT environment. Such techniques will facilitate the identification and extraction of data assumed to have been modified or deleted due to the constraints of IoT devices.

Preservation of the scene is a contentious issue in digital forensics. IoT investigations will complicate matters further due to the nature of the devices undergoing analysis. It is possible that data at a crime scene will be overwritten/compressed if the devices cannot interact with a cloud service provider to store their data, and they collect more data than they can store. This presents a problem for first responders, who must decide whether to preserve the evidence on the devices by allowing data transfer from the scene and then face the challenges of an inter-jurisdiction evidence collection process. Alternatively, they may sever the connection between the devices and the cloud and attempt local extraction of evidence from devices that may be of a proprietary nature. However, the physical placement, power availability or connectivity of each device may render this approach impractical. This also raises the question of whether – and how – a first responder or investigator should prevent devices recording information once a scene has been secured. Principle two of the ACPO guide indicates that a person may access the original data during in an investigation if they are capable of explaining the relevance and implications of

doing so (7safe, 2011). Further research is required to determine what the implications are under a variety of circumstances. Ideally, a mechanism should be in place to enable an investigator to serve a "digital warrant" that prevents evidence being compromised.

We consider the interplay between the legal and technical challenges associated with gathering evidence from an IoT environment. A warrant is served during conventional investigations as the first part of the evidence preservation process. The warrant details the scope of evidence to be seized and examined. In the case of the IoT service providers, they often store data on behalf of their users. This means that individuals may not have direct access to their own data, or it may be presented to them in a different format than that in which it is stored. This complicates the preservation process, as the warrant may have to be served to individuals and their service providers.

### 3.3. Analysis

Analysis of data from an IoT environment will have to consider the provenance of evidence in order to demonstrate the evidence is reliable and authentic. Data provenance in the IoT differs from conventional digital forensic investigations in which the temporal dimension is often the main consideration e.g. file modified, accessed, and created time, email time lines (Inglot, Liu, & Antonopoulos, 2012).

The interaction between IoT and cloud computing facilitates the aggregation and processing of data from the IoT. The vast quantities of data generated by IoT and stored in large-scale distributed cloud environments (Osborne & Slay, 2011) is likely to be the subject of a cloud investigation. From a technical perspective the image, analyse present paradigm of current digital forensics practice (Allen, Whittaker, & Howard, 2005), (Grobler, Louwrens, & von Solms, 2010) does not map well onto the IoT domain. This is aside from the ethical issues of imaging these devices in multi-tenancy cloud environments (Naqvi, Dallons, & Ponsard, 2010), (Burd, Jones, & Seazzu, 2011). There are a number of technical barriers; IoT data is either stored on proprietary devices that are difficult to interface with or in cloud computing platforms where the scale, distribution and remote nature of the data preclude imaging as a viable extraction process. Distributed analysis techniques are required to analyse the data stored in cloud computing platforms. Some work has already been carried out in this area to tackle the challenges posed by cloud computing investigations (Hegarty, Merabti, Shi, & Askwith, 2012), (Garfinkel, 2007).

### 3.4. Presentation

Presenting the findings of IoT investigations poses a new challenge; data will often have undergone aggregation and processing using analytic functions that can alter the structure and meaning of data. At the device level, lossy compression techniques may reduce the granularity of the data order in to preserve limited resources such as memory, battery life, network bandwidth, etc. The granularity and semantics of evidence from the IoT will create challenges to digital forensic investigations. For example, one system may store temperature ranging from 0-5 as cold, 6-10 as average, 11-16 as warm and 16+ as hot. Another system may use different figures

and describe the same temperature readings using different terminology resulting in a semantic gap. Ontological descriptors and standardisation of metadata has limited adoption, with a view to moving IoT devices towards a semantic sensor web (Sheth, Henson, & Sahoo, 2008). However, from a forensics perspective, the issue is that devices may adopt differing descriptor formats or may retain a proprietary format. Presentation poses a challenge regardless of the underlying format of the data, as the conflicting grammar describing data from IoT systems has the potential to be misleading.

# 4. Key Issues and Emerging Requirements

The emergence of the IoT will present new opportunities for data to be misused and lead to an expansion and development of new digital forensic techniques. We identify new approaches that may emerge out of the necessity to analyse the IoT.

## 4.1. Preservation Issues

Firstly we consider the preservation of evidence, forensic readiness is an area that is relatively well understood in conventional computing environments (Pangalos, Ilioudis, & Pagkalos, 2010). We agree with (Trček, Abie, Skomedal, & Starc, 2010) who state that an alternative approach is required to enable IoT forensic environments to achieve the same. As we suggested in Section 3 a digital warrant would assist in the gathering of evidence. This approach could be extended to "digital preservation orders" that prevent evidence from be contaminated or overwritten by reducing the resolution at which data is captured by devices, or freezing the data stored by service providers. The warrant would be digitally signed by the serving authority to enable the providers or devices to check the authenticity of the warrant. The providers or devices would them submit the requested information to the authority over a standard set of interfaces.

## 4.2. Aggregation Issues

The financial motivation behind many IoT systems is the value that comes from the data aggregated in the providers' systems. Such data sets are valuable marketing commodities. Future investigations may benefit from such data to provide or substantiate evidence about an individual or sequence of events.

However, to consider briefly an opposing standpoint; aggregated data may breach data privacy legislation, with the holders of data inferring information about individuals that breaches legislation. New techniques are required to reason over data and determine what can be inferred from large data sets, likewise techniques are required to investigate cases where "aggregation offences" are alleged to have taken place. Similarly, investigatory techniques are required to analyse cases where anonymisation techniques were inadequate, or rendered so by joint analysis of many data sets. Legal frameworks must be updated alongside the development of these techniques to ensure that the data gathered by the IoT is not misused.

While the aggregation of data provides the possibility of inferring useful information about an individual, it also introduces some challenges such as the semantic gap

discussed in Section 3. One approach to tackling this challenge is the development of digital forensic tools that can bridge the semantic gap. These tools would enable calculation and comparison of the granularity of data from different sources. This approach would be particularly useful when conflicting evidence emerges from different IoT devices or service providers. It may be possible to resolve semantic conflicts and even use characteristics of the measurements taken by different systems to provide evidence that is more accurate.

## 5. Conclusion and Further Work

The IoT presents a large-scale source of potential evidence. However due to the heterogeneous nature of the IoT devices, the ways in which data is distributed, aggregated, and processed presents challenges to digital forensics investigations. New techniques are required to overcome these challenges and leverage the architectures and processes employed in IoT to in order to gain access to this rich source of potential evidence.

In order to realise the approaches proposed in this paper, a test bed is required for implementation, deployment, analysis and evaluation. The Contiki open source OS for the IoT ("Contiki OS," 2013) will be used to a test bed. To enable the deployment of techniques on a variety of different devices and network topologies. Experiments will be conducted to evaluate the resource overhead of issuing the digital warranty and preservation orders proposed in this paper. To analyse the trade-off between resource utilisation and evidence gathering, and compare the impact of a variety of certification techniques, that could employed to authenticate the warrants/preservation orders. The test bed will also enable experiments on evidence extraction and the development of standard interfaces for evidence extraction from current and future IoT systems

The development of guidance for investigators on how to carry out investigations in the IoT will be a major output from the development of the test bed. Experimental investigations will enable the identification of considerations that investigators must take into account when investigating the IoT. Along with the development of metrics to determine whether it is better to shutdown devices or leave them running in situ.

Our analysis of IoT forensics has prompted us to consider how we view and deal with data and devices in the broader digital forensics field. Is data a by-product of human-device or device-device interaction, or should the devices be considered as attributes of the data? What are the implications of these two viewpoints for the wider field?

## 6. References

Allen, W. H., Whittaker, E. J. A., & Howard, M. (2005). Computer forensics. *Security & Privacy Magazine, IEEE*, *3*(4), 59–62. doi:10.1108/09565690610677463

Attwood, A., Merabti, M., & Abuelmaatti, O. (2011). IoMANETs: Mobility architecture for wireless M2M networks. In *2011 IEEE GLOBECOM Workshops (GC Wkshps)* (pp. 399–404). IEEE. doi:10.1109/GLOCOMW.2011.6162479

Burd, S. D., Jones, D. E., & Seazzu, a F. (2011). Bridging Differences in Digital Forensics for Law Enforcement and National Security. In *2011 44th Hawaii International Conference on System Sciences* (pp. 1–6). Manoa: Ieee. doi:10.1109/HICSS.2011.87

Chirgwin, R. (2014). SPAM supposedly spotted leaving the fridge. *The Register*. Retrieved from http://www.theregister.co.uk/2014/01/20/spam_spotted_leaving_the_fridge/

Contiki OS. (2013).

Garfinkel, S. S. (2007). Commodity grid computing with Amazon's S3 and EC2. *Usenix*, 7–13.

GM Casualty Reduction Partnership. (2014). Mobile Phones | Greater Manchester Casualty Reduction Partnership.

Grobler, C. P., Louwrens, C. P., & von Solms, S. H. (2010). A Multi-component View of Digital Forensics. In *2010 International Conference on Availability, Reliability and Security* (pp. 647–652). IEEE. doi:10.1109/ARES.2010.61

Haines, L. (2007). Cops may check crash drivers' mobile records • The Register. *The Register*. Retrieved from http://www.theregister.co.uk/2007/02/27/mobile_phone_proposal/

Hegarty, R., Merabti, M., Shi, Q., & Askwith, R. (2012). Scalable Distributed Signature Detection. In *Proceedings of the 7th International Workshop on Digital Forensics & Incident Analysis* (pp. 27 – 37). Heraklion, Greece.

Kent, A. K., Chevalier, S., Grance, T., Dang, H., & Kent, K. (2006). Guide to integrating forensic techniques into incident response. *NIST Special Publication*, (August).

McKemmish, R. (1999). What is forensic computing. *Trends and Issues in Crime and Criminal Justice*, (118).

Miorandi, D., Sicari, S., De Pellegrini, F., & Chlamtac, I. (2012). Internet of things: Vision, applications and research challenges. *Ad Hoc Networks*, *10*(7), 1497–1516. doi:10.1016/j.adhoc.2012.02.016

Naqvi, S., Dallons, G., & Ponsard, C. (2010). Applying Digital Forensics in the Future Internet Enterprise Systems - European SME's Perspective. In *2010 Fifth IEEE International Workshop on Systematic Approaches to Digital Forensic Engineering* (pp. 89–93). IEEE. doi:10.1109/SADFE.2010.28

National BIM Standards. (2013). Retrieved from http://www.nationalbimstandard.org/faq.php#faq1

Oriwoh, E., Jazani, D., Epiphaniou, G., & Sant, P. (2013). Internet of Things Forensics: Challenges and Approaches. *Proceedings of the 9th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing*. doi:10.4108/icst.collaboratecom.2013.254159

Osborne, G., & Slay, J. (2011). Digital Forensics Infovis: An Implementation of a Process for Visualisation of Digital Evidence. *2011 Sixth International Conference on Availability, Reliability and Security*, 196–201. doi:10.1109/ARES.2011.36

Pangalos, G., Ilioudis, C., & Pagkalos, I. (2010). The Importance of Corporate Forensic Readiness in the Information Security Framework. In *2010 19th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises* (pp. 12–16). IEEE. doi:10.1109/WETICE.2010.57

Sheth, A., Henson, C., & Sahoo, S. S. (2008). Semantic Sensor Web. *IEEE Internet Computing*, *12*(4), 78–83. doi:10.1109/MIC.2008.87

Taylor, M., Haggerty, J., Gresty, D., & Hegarty, R. (2010). Digital evidence in cloud computing systems. *Elsevier Computer Law & Security Review*, *26*(3), 304–308. doi:10.1016/j.clsr.2010.03.002

Thomas, S., Sherly, K. K., & Dija, S. (2013). Extraction of memory forensic artifacts from windows 7 RAM image. In *2013 IEEE CONFERENCE ON INFORMATION AND COMMUNICATION TECHNOLOGIES* (pp. 937–942). IEEE. doi:10.1109/CICT.2013.6558230

Trček, D., Abie, H., Skomedal, A., & Starc, I. (2010). Advanced framework for digital forensic technologies and procedures. *Journal of Forensic Sciences*, *55*(6), 1471–80. doi:10.1111/j.1556-4029.2010.01528.x

Zelkha, E., Epstein, B., Birrell, S., & Dodswoth, C. (1998). From Devices to "Ambient Intelligence." In *Digital Living Room Conference*.