Digital evidence in fog computing systems

R Hegarty,
Independent Computer Security Consultant,
Manchester,
UK
Email: rob@hegarty.info

M Taylor*
Department of Computer Science and Mathematics,
Liverpool John Moores University,
Liverpool,
L3 3AF
UK
Email: m.j.taylor@ljmu.ac.uk

*Corresponding author

Abstract

Fog Computing provides a myriad of potential societal benefits: personalised healthcare, smart cities, automated vehicles, Industry 4.0, to name just a few. The highly dynamic and complex nature of Fog Computing with its low latency communication networks connecting sensors, devices and actuators facilitates ambient computing at scales previously unimaginable. The combination of Machine Learning, Data Mining, and the Internet of Things, supports endless innovation in our data driven society. Fog computing incurs new threats to security and privacy since these become more difficult when there are an increased number of connected devices, and such devices (for example sensors) typically have limited capacity for in-built security. For law enforcement agencies, the existing models for digital forensic investigations are ill suited to the emerging fog paradigm. In this paper we examine the procedural, technical, legal, and geopolitical challenges associated with digital forensic investigations in Fog Computing. We highlight areas that require further development, and posit a framework to stimulate further consideration and discussion around the challenges associated with extracting digital evidence from Fog Computing systems.

Keywords: digital evidence, fog computing, cyber crime

1. Introduction

The term fog computing was introduced in 2012 by Cisco for dispersed cloud infrastructures. Fog computing provides the ability to handle a large number of Internet of Things devices and big data volumes for real-time low-latency (minimal delay) applications (Bonomi et al., 2012). Fog computing extends the cloud computing model in terms of low latency, location awareness, very large numbers of heterogeneous nodes, predominantly wireless access, and real time exchange of sensor data. The fog computing model incorporates Internet of Things (IoT) services and applications (Sullivan, 2019), such as connected vehicles, smart cities, and more generally, wireless sensor and actuators networks (WSANs) (Bonomi et al., 2012). In this paper we discuss a framework for fog forensics that examines the scope of the digital investigation, the time frame of the investigation, the nature of the fog system being investigated, the methods of data capture, and level of access to fog system components. Fog computing is different to cloud computing and Internet of Things configurations, since fog computing involves the use of a number of computing resources for processing and the use of sensor and actuator devices, that is it combines the two paradigms, and is therefore an appropriate and useful classification for forensic techniques. Digital investigations of fog
computing systems may be complicated due to the different types of devices that they contain, these devices may be part of private and public networks. The originality of the research presented in this paper is the development of a framework specifically for fog forensic investigations.

2. Fog Technology

Fog computing provides a new approach to the distributed storage and processing of data. The architecture of fog computing systems can include conventional client server devices, as well as Internet of Things (IoT) devices, connected vehicles, and wireless sensors. While the focus of cloud technology has been the provision of elastic computing and storage resources in the core of the network to facilitate a wide range of “as a Service” models IaaS, PaaS and SaaS (Infrastructure, Platform, Software as a Service respectively). Fog computing focuses on a broader range of technologies, encompassing; Cloud, IoT, Industry 4.0. The platform and service requirements of Fog computing typically include low latency machine to machine (peer to peer) communication at or near the network edge. This is supported via cloud infrastructure to provide long term storage, accounting, and oversight. Figure 1. illustrates the relationship between cloud computing, the Internet of Things and Fog computing.

![Figure 1 Fog Computing model](image)

3. Legislation relevant to fog computing forensics

Under English law, a computer hard disk is defined as a single storage entity and is considered a document (Mason, 2018) since it is something ‘in which information of any kind is recorded’. For fog computing systems, forensic imaging of data from all of the computing devices (or even a subset of the devices) and sensor devices in a fog system may not be practicable, especially in terms of the seizure of the devices (HO, 2016). An investigation into a fog system should only search ‘places which might reasonably be suspected
of containing the specified offending articles’ (Stone, 2013). Investigating a fog computing system could make targeting a single user without capturing data from other users almost impossible. This complication blurs the line between mass surveillance and targeted law enforcement investigations. Fog computing service providers may record certain information relating to use of their services, however storing data may impact the latency of low powered devices. To maintain the requisite latency and performance, some devices within a fog computing system are unlikely to be logging much data.

In the absence of evidence to the contrary, the courts will presume that a computer is in working order at the material time. An important aspect of providing digital evidence in court concerns certifying that the computer(s) in question were working properly at the material time (Mason and Seng, 2017a; Marshal, 2020). A person in physical control of a computer or other device in a fog system would typically not be aware whether it is working ‘reliably’, ‘properly’, ‘consistently’, ‘correctly’ or ‘dependably’ (Mason and Seng, 2017b). The UK Post Office Horizon scandal has brought the issue to the fore (Mason and Seng, 2017c; Ladkin et al., 2020; Christie, 2020; Marshal et al., 2021). Given the scope and scale of fog platforms, and considerations around the robustness of all the software involved in such (Ladkin, 2020) it is possible that some software within the fog system may not have been in working order at the material time.

With regard to the legal liability of fog service providers, there would not be liability for criminal offences resulting from hosting data or applications, provided that there was no actual knowledge of unlawful activity, and there was no reason to suspect such unlawful activity. If end-to-end encryption is applied in a fog computing system, the fog service provider does not hold the encryption keys and would not be able to view the data, and thus could not be held liable, provided that they were unaware that and had no reason to suspect that any illegal content or activities were occurring, as per the EU E-commerce Directive 2000. Data may be encrypted within the user’s computer prior to transmission within the fog system. Even if users intend to process data unencrypted in the fog system, the system may be set up to encrypt all or part of the data it receives (Hon et al., 2011). The person having possession of information or a key to protected information in a fog system, is defined in s 56(2), RIPA 2000 (Mason and Seng, 2013d). It is a criminal offence knowingly to refuse or fail to make the disclosure required by a notice issued under s 49.1 of RIPA 2000 (Mason and Seng, 2013e). In the UK, the GCHQ’s ghost proposal for law enforcement agencies to act as a silent added participant in private communications could potentially make law enforcement investigations of fog computing applications and services easier. Using the ghost proposal, an investigators device could be added to a fog network, without the knowledge of participants, from here it would be able to capture and decrypt evidence from the other devices.

In the UK, the Data Protection Act, 2018, and the European Union General Data Protection Regulation (2016) provides security for personal data accessed, however other jurisdictions may not have a similar level of legislation (Sullivan, 2019). In particular, Privacy shield is an agreement between the European Union and the United States that concerns the transfer of personal data from the European Union to the United States. The European Union General Data Protection Regulation (2016) (GDPR) has specific requirements regarding the transfer of data out of the European Union, requiring that the transfer must only happen to countries deemed as having adequate data protection laws. In general, the European Union does not list the United States as one of the countries that meets this requirement. Privacy Shield was designed to create a program whereby participating companies would be deemed to have adequate protection, facilitating the transfer of information. Privacy Shield has recently been invalidated by the Schrems II case, which found the protections for data transfer between the EU and US were inadequate. In more general terms, ongoing geopolitical change may affect compliance, for example Brexit which in some instances may result in the UK being treated as a third party. Although such geopolitical changes currently happen infrequently, as the value of data continues to increase for organizations, countries may have different priorities around privacy and data sharing.

The UK Criminal Procedure and Investigations Act, 1996 (CPIA) and amendments in the UK Criminal Justice Act, 2003 (Part 5) (CJA) may be relevant with regard to providing both evidence in support of a prosecution and evidence to support a reasonable defence in fog computing forensic investigations (e.g. R. v. Hampton and another 2004 EWCA Crim 2139, concerning an example case where non-disclosure of cell-
site evidence relating to a mobile phone call occurred). The UK Regulation of Investigatory Powers Act, 2000 (RIPA, 2000) could apply to the monitoring of fog computing systems during a computer forensic investigation in certain organisations, due the requirement upon organisations to support police investigations.

The UK Computer Misuse Act 1990 (and subsequent legislation) offences may prove difficult to investigate due to acquiring relevant data from fog computing systems (e.g. what fog computing users accessed and manipulated which fog based resources). For criminal procedure, a fundamental balance needs to be found between the rights of the suspect, in terms of privacy and fair trial, as well as other individuals collaterally involved in, or impacted by, an investigation, and the needs of law enforcement to investigate and prosecute offenders (Walden, 2016). Any search for evidence must be restricted to the extent necessary to achieve the objective of the search. Searches must be conducted with due consideration for the privacy of the detained person or occupier (HO, 2016).

Money laundering and fraud-based investigations (involving money laundering legislation) might have to rely on audit trails created by applications (especially accounts systems). These audit trails may be further complicated through the use or crypto-currencies, which while readily traceable (Fleder et al, 2019), presents challenges around the anonymity of the individuals responsible for payments.

### 4. Challenges and opportunities within fog computing forensics

**Challenges within fog computing forensics**

Wang et al (2015) commented upon the issues and challenges in security and forensics for fog-based systems. Data stored and processed in fog-based systems may be in countries where privacy laws are not particularly strong, readily enforced, or non-existent. The European Union has strong privacy laws such as GDPR (Sullivan, 2019), and there are increasing calls for United States privacy laws to catch up, in particular in terms of increased regulatory control of the big data companies such as Google, Amazon, and Facebook. Individual states are already implementing stronger laws, making the emergence of federal legislation inevitable.

Access to data in fog-based systems prior to it being seized, and preservation of such data is an important issue, since due to dynamic nature of the operation of fog computing systems, it would not be possible to go back to the original state of the data. Data collection within fog-based systems could impact latency, and in general data collection is complicated by the ephemeral nature of data in a fog domain. It is becoming increasingly challenging to determine what type of data should be collected from IoT devices and how traces from such devices can be leveraged by forensic investigators (Al-Masri et al, 2018). When undertaking forensic investigations of large fog-based systems there can be computational load issues associated with large-scale data set searches of fog-based systems. The issue of what constitutes data and evidence is further complicated by the process from which information is inferred and deduced from data in fog systems. Often the raw data itself is discarded, with only the inferred or deduced information stored. Relying on such information, would introduce the requirement for investigators to have access to the proprietary machine learning algorithms used make inferences and data sets used to train them. The algorithms are often trade-secrets and the datasets used to train them contain data from many hundreds of thousands of individuals. Here again investigations face the challenge of being too broad in scope, and thus considered instruments of mass surveillance. These practical issues aside, the complexity of the algorithms and systems make developing a detailed understanding impractical for many investigators.

Confiscating physical computing equipment as part of a computer forensic investigation could be difficult in a fog computing environment due to geographical dispersion, and some computing devices potentially being in different jurisdictions. There may also be safety critical and contractual obligations to consider. For example, if the fog network around a hospital system were seized there could be patient safety issues, and potential issues concerning service level agreements and associated financial penalties.
Patrascu and Patriciu (2013) commented that forensic investigators need to avoid copying entire disk images. In fog and cloud environments, data sizes can be very large and stored across multiple clusters, each containing hundreds of disks. In such cases it will not always be technically possible or cost effective to attempt to copy the relevant storage device images. Thus, there need to be accepted “forensically sound” mechanisms whereby instead of copying an entire disk image, only the data (e.g. Device Settings, Memory Dumps, Packet Captures, Database Entries) representing evidence would need to be copied. Further to this, data in fog computing systems is often fragmented and duplicated across many proprietary low power devices, making low-level image capture a non-trivial undertaking (Al-Masri et al, 2018). An accepted forensically sound mechanism for copying the data representing evidence might have to take the form of a standardised application programming interface (APIs), web services or physical JTAG (Joint Test Action) connections.

Overall, currently there does not appear to be an established method for extracting digital evidence in an admissible fashion from fog-based applications.

Acquisition of digital evidence in fog computing systems

The manner of fog computing service operations means that in practice, an organization may not know where data for which it is responsible geographically located at any given time. Before considering the legal and procedural approaches to capturing evidence from fog computing systems, the types of evidence must first be classified. The table below provides a taxonomy of evidence types and potential approaches to processing evidence.

<table>
<thead>
<tr>
<th>Evidence</th>
<th>Location</th>
<th>Acquisition Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encryption Keys</td>
<td>Memory of IoT device</td>
<td>JTAG Memory Dump</td>
</tr>
<tr>
<td>Network Packets</td>
<td>Fog network infrastructure</td>
<td>Packet capture via network tap</td>
</tr>
<tr>
<td>Metadata</td>
<td>IoT devices and network traffic</td>
<td>Device firmware, packet capture</td>
</tr>
<tr>
<td>Malware binaries</td>
<td>Memory of IoT devices, Malicious firmware</td>
<td>JTAG Memory Dump</td>
</tr>
<tr>
<td>User data</td>
<td>Cloud storage, IoT devices</td>
<td>API interface, web service, device or disk imaging via JTAG</td>
</tr>
</tbody>
</table>

Table 1 Summary of evidence types in fog computing forensic investigations

The extracting of encryption keys from IoT devices, is challenging due to the proprietary nature of both the software/firmware and physical connections to these devices. The capture of network traffic is a well-established approach for gathering evidence, however the emergence of end-to-end encryption makes this process untenable in the long term. Malware is increasingly becoming diskless to avoid detection via conventional methods; memory analysis of proprietary platforms brings about the challenges described previously. User data while dispersed by the nature of fog platforms, may be the most directly accessible form of evidence, through either an API or web service. Such web services and API’s will require strict access control, as they present an obvious attack surface. The figure below provides a high-level illustration of the various inspection points required to analyse evidence in fog computing system used to facilitate smart transportation. As is typical with many fog computing scenarios, the collection of evidence is...
complicated by the mobility of devices, and complex supply chain of cloud providers, network providers, vehicle manufactures, infrastructure (traffic signal) controlled by various stakeholders, etc.

In a fog computing model devices may move into and out of the fog network continually. In order to create resilience and persistence, fog computing devices may receive copies of data from other devices, despite not being part of the network at the material time. This and the abstract nature of fog computing systems could create more complex chain of evidence scenarios for forensic fog computing investigations.

![Diagram](image)

**Figure 2. Analysis of digital evidence in fog computing systems**

Figure 2 depicts the different sources of digital evidence in a smart traffic fog system, where digital evidence could be gained from vehicles, traffic light controllers, mobile telephone masts, and cloud based servers. Meta data (for example, geolocation via point of connection, or mac address) may prove useful to forensic fog computing investigations in terms of determining how data stored in a fog computing system has been used and manipulated. It may also serve as evidence of the geolocation of devices in the physical world. For example, a connected stolen car may pass traffic lights, and leaves behind evidence in the form of metadata (it’s MAC address for example).

Traditionally, email clients and email servers within a computer systems contain logs of sent and received emails useful for investigating email correspondence, however, the fog computing approach breaks the client server model, with services being offered using more distributed mesh type protocols. This means that investigating what would traditionally be deemed single user client devices, may now be deemed investigation of hybrid client / server devices storing evidence from multiple users of a service. This poses challenges to the established norms around completeness of evidence, and the scope of warrants. Furthermore, the skill set required to investigating the ever increasing complexity of such emerging systems, will required support and training for the investigators.

Investigations concerning malware (such as virus, worms and spyware) within a fog computing environment may be complex, since detecting the presence and effects of such malware upon data or programs stored within the fog computing could be rather complex. Alasmary et al (2019) stated that the steady growth in the number of deployed Internet of Things (IoT) devices has been paralleled with an equal growth in malicious software (malware) targeting such devices. A defence related to malware within fog-based system unknown to the accused, may have difficulty in obtaining appropriate digital evidence to support such a defence. Typically, fog-based systems are likely to need to run bespoke intrusion detection systems, as the constrained devices will likely be unable to protect themselves. Such intrusion detection
systems could store and aggregate potential evidence. Furthermore, Fog networks may be the target or origin of disruptive Denial of Service attacks; for example, the Mirai botnet of IoT devices (C. Kolias, et al, 2017). Defence against such attacks typically requires filtering by a third party. In the case of Mirai both Akami and Google provides such services. These filtering services store and process huge amounts of potential evidence. While these services have an established track record of collaborating with law enforcement to support prosecution of international organised crime networks, the patchwork of national and international data privacy legislation may make the investigation of smaller cases more difficult.

*Procedures for fog computing forensic investigations*

Fog computing forensics is distinct from just dealing with each device as either an IoT device or cloud computing provider in that it requires an integrated approach that considers all devices from sensors to all the processing devices involved in a given operational fog system activity. That is, it requires a holistic forensic approach that caters for all devices and communication channels that may be involved in the fog computing system when a given operational transaction or activity is undertaken.

A procedure for fog computing forensic investigations would initially involve identification of data of interest within the fog computing systems concerned. A computer forensic investigation involving a fog computing system should ideally not impact upon other fog service users who are not the target of the investigation. A procedure for fog computing forensic investigations would also need to comply with relevant police guidelines in the relevant jurisdiction (e.g. the ACPO guidelines in the UK). Below we outline our framework which augments the CFSAP Computer Forensics, Secure, Analyse, Present (George et al, 2003) forensic response model to account for the specialist nature of fog computing investigations.

**Device Identification and Scoping**

During this stage the scope of the investigation is used to determine how many devices within the fog system need to be included, furthermore the time period from which the evidence is gathered is also identified. Finally, the time constraints on capturing and securing the data are investigated. In some instances, data is already preserved in immutable storage, in others it may be more ephemeral and require immediate capture.

**Composition**

This stage involves detailed examination of the composition of the fog system being investigated. Of the IoT devices to be involved in the investigation, how many just transmit data (e.g. a pure sensor), how many have stored code, and how many have stored code and data.

**Methodology**

This concerns determining the appropriate methods of data capture from the different devices within the fog system. The IoT devices to be involved in the investigation, can then be easily and safely removed. For example, an IoT device might be integrated into a complex production machine and might not easily be removed. As another example, an IoT device might be part of a safety critical system, such as life support equipment in a hospital, and might not safely be removed due to operational requirements.

**Implementation**

The final stage involves examining in detail the level of access to the fog system components. The nature of extraction of data and code contained in an IoT device will depend upon the make / model of the IoT device. In practical terms, the investigator will require a full set of devices and tools with which to remove IoT devices (if appropriate) or extract data and code from such devices. In some instances, collection of
data for the investigation may require physical interception of data transmitted via cabling, and wireless transmission of data transmitted via a wireless network component of the fog system.

Each of these phases of the framework extends the CFSAP model by augmenting the Secure and Analysis phases. During the device identification and scoping phase of the framework, a specialised extension of the secure phase of the CFSAP model is proposed, that considers the scale, complexity and dynamic nature of fog investigations. During the composition phase, the secure and analysis phases of the CFSAP model are enhanced to enable the implications of the structure of the fog system to be considered. The methodology phase of the framework takes account of the complex real-time nature of the cyber physical systems found in cloud computing, and the implications to the investigation. The implementation stage of the framework refines the secure and analysis phases of the CFSAP model, by taking account of the specialist tools, techniques and access rights, required to analyse a fog based evidence.

5. Case Study
In order to investigate the forensic challenges posed by fog computing and develop a framework to highlight or address these challenges a prototype fog system was deployed. The system gathers environmental data from two Nordic nRF6936 IoT Sensor kit devices (https://www.nordicsemi.com/Software-and-Tools/Prototyping-platforms/Nordic-Thingy-52). The Nordic devices run an embedded operating system on a system on chip that gathers data from the device’s onboard sensors (Temperature, Pressure, Humidity, Volatile Organic Compounds, etc), this data is exposed via a Bluetooth 5.0 API for transfer to a staging device before being processed and visualised in the cloud.

The staging device employed in our prototype is a Raspberry Pi Zero W single board computer. This device aggregates local data before forwarding it to a web service. Web services are typically hosted in the cloud. However, our prototype uses a second more powerful Raspberry Pi 4 single board computer as a cloud proxy. This device hosts a web service, web server and basic visualisation platform to enable end users to visualise trends in the environmental sensor data.

![Figure 3 Prototype Platform](image-url)
Throughout the case study, we highlight the considerations associated with extracting data from a larger commercial fog platform. Below we describe the application of our framework.

Device Identification and Scoping

Identification of the devices in our case study was straightforward, as both the types and locations of the devices were known. In a large-scale commercial fog system, coordination with the owner/provider of the system would be required to gather the detailed architectural documentation required to locate sensors, aggregators and cloud systems. If such coordination is not feasible, because for example the owner or provider is a suspect in the investigation. Wireless spectrum analysis would be required to locate sensor devices, the feasibility of such an undertaking will be limited if either a large geographic area is covered by the network, or the network contains a large number of sensors. Similarly, without coordination, network monitoring techniques would be required to locate the IP addresses and services providers associated with the cloud systems responsible for hosting the fog network services and data.

Determining the duration required to capture data, and the duration in time from which the data is captured, was once again trivial for our case study, having access to months’ worth of data through the end user facing interface, and raw data on the staging device and cloud device. The scale of data stored and processed by a large-scale commercial fog system, would likely bring a financial dimension to an investigation. The quantity of data stored in such systems, and the computational power required to analyse such data, will inevitably introduce a cost / duration trade-off for the investigator.

Composition

While the composition of our case study system was limited. The complexity heterogeneity of devices in even this small-scale system threw up some challenges. The sensor devices where accessible via an API, I/O expander and programming/debugging connector, we were able to retrieve live data from the sensor, however this interrupted the service provided to the aggregator and resulted in data loss occurring on the fog system. The staging device made use of an open-source Bluetooth Low Energy library to poll the sensors. This device used a username and password combination known to us. We were able to swap out the device by cloning its storage and booting a replacement device with the cloned image. Unfortunately, this resulted in a keying mismatch, the prevented the staging device from communicating with the cloud device, once again resulting in service interruption and data loss on the fog system. In a commercial system, the API's required to gather the data required for an investigation are often not publicly accessible.
Requiring special permission and support from the infrastructure provider to gain access. The scale of a commercial fog system is also much larger, making determining the composition of the system more challenging. In some instances, the exact composition of the system may be undocumented or considered a trade secret.

This case study demonstrates the practicalities of fog computing forensics. Personal data would be included if the user could be identified from identification codes included in the system. Clearly, in more complex fog computing systems such as smart homes there would be numerous devices, however, the practical forensic aspects would therefore just be multiplied by the number of devices and their different configurations. Similarly, larger amounts of data in a more complex fog computing system would involve much more time and effort to search for evidence. Uncooperative service providers and international jurisdictions would be dealt with on a case by case basis, relating to the authority of the police or other agency.

Methodology

The methodology used to extract data from our case study mirrored the techniques used in the development of the system. Such an approach assumed that access will be granted with the system owner/designers’ consent, and that support will be available to understand how the system was created and how it functions. Given the dynamic nature of such complex systems. It is likely that gathering the required documentation and know how to understand a commercial fog system, will be a significant undertaking from both a bureaucratic and technical perspective.

Implementation

The implementation employed in our case study was predicated both in-depth knowledge of the architecture, API’s and source code employed by the system. Even this knowledge and unrestricted access, our investigation triggered data loss in the system. This highlights the challenge of conducting live digital forensic investigations in a dynamic networked environment such a fog computing. In a commercial system, many of the protocols, APIs and much of the source code associated with the system may be beyond the reach of the investigator. This will likely make gathering complete evidence from such a system infeasible without the support of the system owner and designer.

6. Conclusions and Future Work

Acquiring and analysing digital evidence from fog computing systems is likely to be more complex than for other types of systems. Unless a fog-based application provides audit trail data, it may be difficult to extract admissible digital evidence from such applications.

The nature of data and the processing used to infer and deduct information will require a step change in cooperation between providers and law enforcement, accompanied by additional training to empower investigators to understand the concepts behind the powerful inference engines employed in the fog domain.

Increased requirements for data portability in GDPR have strengthened the requirement for audit trails and the right to be forgotten in fog computing systems (Rawat, et al, 2019), whilst the right to be forgotten appears to be in conflict with digital forensics investigations. Ideally, legislation could mandate privacy preserving mechanisms for data collection from specific individuals within fog computing systems. As newer more complex fog-based systems emerge this is likely to become more of an issue for computer forensic investigations.

References


R. v. Hampton and another 2004 EWCA Crim 2139


