Abertay University

UNIT 2

[Redacted]

CMP416: Advanced Digital Forensics

2024/25

Note that Information contained in this document is for educational purposes.

.

Contents

.

1 INTRODUCTION

1.1 BACKGROUND

Digital forensics is the process by which data stored digitally in a wide range of sources such as computers or smartphones is analysed and interpreted, typically with the purpose of investigating a crime or event that took place. Evidence of this type occurs in essentially every type of crime in the modern age, which is no surprise when it was estimated in 2020 that for every person on earth, 1.7 MB of data was created every second (DOMO, 2018). This makes clear that digital forensics is an important tool for investigators, with it being ever more prescient in crimes that occur exclusively in the digital world where the amount of data left behind is more significant with digital forensics often being the only way for any evidence to be identified.

IOT (Internet of things) devices refer to devices connected via a network that allow for the exchange of data, often for user convenience. At a consumer level, these devices are typically Wi-Fi (802.11) enabled and have a singular or limited purpose – for instance, lightbulbs, thermostats or kettles, which are often called "Smart" devices. Industry also makes use of IoT devices, with sectors such as manufacturing taking advantage of wireless sensor data. This in turn means that IoT devices represent a significant risk factor, given they often lack the resources to have robust individual security due to their simplicity and there can be a significant number of them, often with differing software on the same network increasing the propensity for security loopholes. This may go some way to explain why IoT malware attacks increased 400% in the first half of 2023 compared to 2022, with 54.4% of those attacks being targeted at the manufacturing sector (Gandhi, 2023). This clearly shows that IoT devices represent a significant breach risk and are increasingly targets of malicious attackers, with this number likely to go up as the number of devices increases as it has been shown to in recent years (Figure 1).

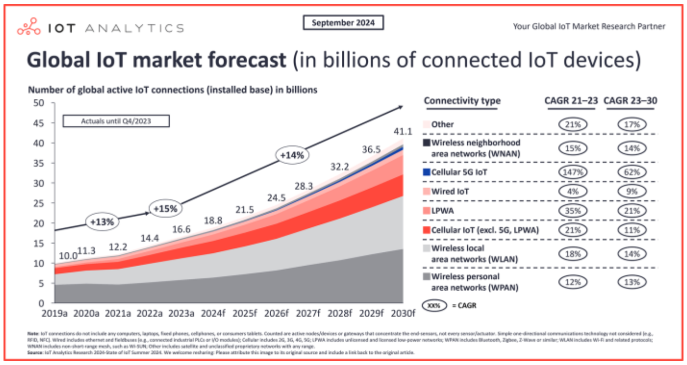


Figure 1 - Predicted growth of IoT Devices 2024-2030 (Sinha, 2024)

To this end, the investigator has been provided with a scenario in which there has been a breach in a smart home featuring a number of consumer grade IoT devices, as a part of a "Smart city" system. The goal is to describe the methods through which a digital forensic investigation would be undertaken, demonstrating elements such as how the data would be obtained, the tools used and the methods of identifying unauthorized access.

1.2 Aim

The aim of this project is to write a research report based on the provided scenario to indicate what actions would be taken by a digital forensic investigator in order to identify evidence of compromise in a forensically sound and appropriate manner within a theoretical IOT network. A number of sub aims encompass this:

- Determine the scope and source of the breach
- Describe the data sources necessary for collection and methods of analysis
- Detail strategies to Identify evidence of unauthorized access and the tools used to do so

2 ACQUISITION & INVESTIGATION STRATEGY

2.1 ANTICIPATED STRATEGIC FLOW

The strategic flow of the investigation will be employing an investigative methodology that is based on the OCF: Open cloud forensics model with relevant modifications to best fit the scenario (Zawoad, et al., 2015). This was chosen due to the fact that it closely matches the methodology established in the coursework and the research paper was based on an actual civil lawsuit, proving it is fit for purpose for this kind of investigation. The report will follow this structure.

The methodology is comprised of these steps:

- Identification – Identify the purpose, scope and resources required for the investigation.
- Preservation – Isolate, secure and preserve the data.
- Analysis – Process the data using relevant tools and techniques, then interpret the results.
- Documentation – Mapping the crime scene, in this case the network topology present.
- Presentation – Summary and conclusion based on the obtained information.

This can be seen again below in the systematic diagram (Figure 2).

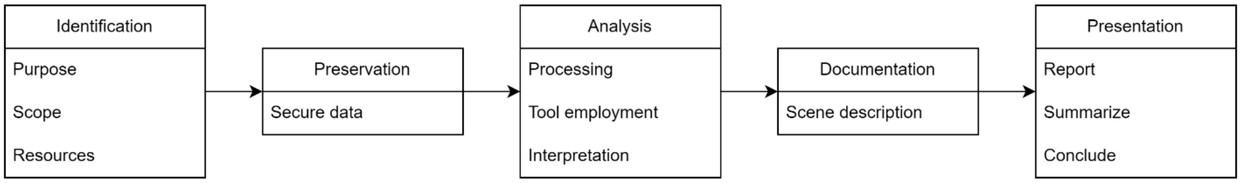


Figure 2 - Systematic Diagram detailing methodology employed

2.2 IDENTIFICATION

2.2.1 Purpose

The purpose of the investigation is to identify the manner of exploitation that lead to the breach of the smart home detailed in the scenario. This will require identifying the impacted devices, how data will be recovered from them and what that data will be – and then how that data will be analyzed to successfully achieve this goal.

2.2.2 Resources

The resources required for this investigation are likely to be fairly extensive due to the variety of devices comprising the network topology. This is due in large part to potentially specialized equipment that may be required for the obtaining of local data on proprietary systems. For example, while obtaining forensic images of computer systems or IOT devices such as the Rasberry Pi is fairly trivial – obtaining forensic images of Smart televisions or thermostats may represent a significant problem and the files obtained may be proprietary requiring reverse engineering efforts to obtain relevant data. This is compounded by the fact that some network devices may use RTOS (Real time operating systems) such as the prevalent ESP32 microcontroller – this means that data is usually not stored, making local data recovery unlikely.

On account of the fact that the system is IOT – the primary source of evidence will likely be located in event and network logs. This should not require significant resources as a number of excellent open source digital forensics tools exist that allow for analysis of successfully logged data providing it is present. A single investigator can likely perform this with a single computer.

2.2.3 Scope

At present the scope of the scenario appears limited to the known breach, which relates to a single smart home. This includes all elements in the individual smart home, including perimeter firewalls which will likely prove instrumental in identifying the entry point.

This does not mean to say that throughout investigation the scope may increase, as while a firewall was in place protecting the centralized system, there is no reason to believe a sufficiently competent attacker would not be able to bypass such an obstacle meaning that fundamentally the scope of the investigation could expand to encompass the whole smart city network if reason to suspect compromise of the centralized server exists. The present elements likely to be forensically investigated as potential areas that contain relevant data are:

- Smart Thermostat
- Smart TV
- Smart Camera System
- Wireless Router
- “Video System”
- Perimeter firewalls
- (potentially) Connected mobile phone

2.2.4 Security Weaknesses/Breach vectors

Based on the information provided in the scenario there are practically only two entrypoints for data transmission to the breached smart home network:

- External connection from the internet
- Internal transfer from centralised network

Both of these would require an attacker to theoretically traverse a firewall, though depending on how the firewall is configured this could have been trivial (for instance, the firewall connected to the centralized server may be significantly weaker to allow for easier engineer access). Furthermore, while a comprehensive list of devices is not present, many smart TV’s allow for comparative internet access to a computer and may even allow for download of files from which an attacker could pivot to fully accessing the other devices. This is relative to thermostats and security cameras which may only download files from specific sources in automatic OTA (Over the air) updates, making them a less likely method of exploit. It is also fair to say that the significant majority of breaches originate from an external internet source, such as viruses, exploited misconfiguration etc. – making this by far in a way the most likely entry point.

An internal breach would be much less likely but much more significant as it would imply that the attacker already has access to a point from which they could access other smart home systems. This is worth consideration but should be treated as a less likely option.

2.2.5 Scenario modelling

An attempt was made to create a diagram illustrating the rough topology of the network as described in the scenario (see Figure 3). This allowed for easier visualisation of the data flow that could aid in explaining the attack flow to city officials. This is based on the information we know:

- Individual internet link per home
- Perimeter security devices per home, meaning firewalls are likely on the exterior with no DMZ.
- Connected to a centralized system, hence "smart city"
- A switch is not mentioned, and a home IOT network would likely use the router.

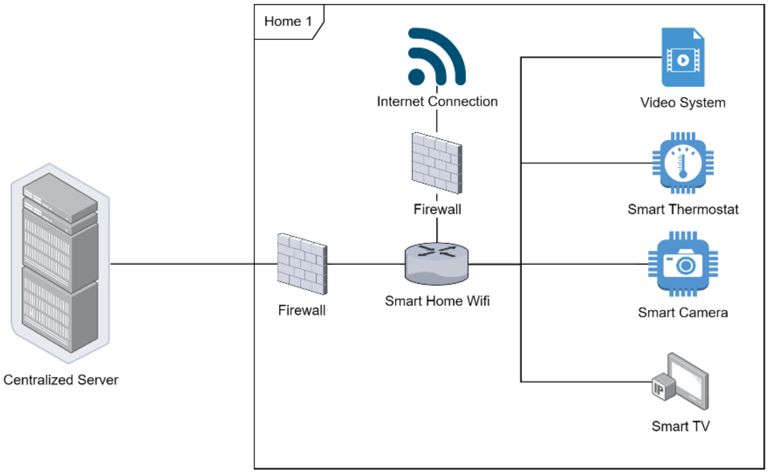


Figure 3 - Attempted reconstruction of network topology

2.3 Preservation of Data Sources

2.3.1 Forensic Image obtainment

In order to preserve the data, forensic images of all the devices should ideally be obtained. This is due to the fact that it allows for the consistency and reproducibility of data – as otherwise it is likely in the process of analysis the investigator may overwrite or otherwise contaminate evidence – making it useless. This is especially pertinent in that many IOT devices have small amounts of memory that will get overwritten with heavy command usage, as in the case of an investigator audit. In many cases these will store data in DRAM (Dynamic RAM), meaning turning off the device should be postponed until a forensic image can be obtained as otherwise data will be lost.

This can be done in a variety of ways and will depend on the device and operating system. For instance, if the device runs Linux it is likely that a user can SSH into it or use some other form of remote CLI, and then produce a forensic image using the "dd" tool in Linux. Other more proprietary methods may require the user to physically access the router to obtain a forensic image such as a JTAG connection or GPIO (general-purpose input/output) (Findlay, 2021).

Often times Smart Devices will have companion apps on a smartphone or other device, and this data can be pulled from those applications. In the absence of this, screenshots may be the most effective method of obtaining “forensic” data, though this requires an immense degree of caution due to the aforementioned overwriting of memory/data and may not stand up in court if not carefully logged.

2.3.2 IoT Devices (Thermostat, Smart TV, Smart Camera system, Video system)

These devices are counted under the same preservation step, as while the method for obtaining forensic data from them may differ – the relevant data obtained is likely to remain the same on account of the fact they are all Wi-Fi (802.11) enabled devices. While it is true that unique data can likely be found on all of these due to their type, for example:

- Temperature data from the thermostat
- Camera footage from the camera system
- Activity data indicating watch time on the Smart TV

These are unlikely to relate to the breach and should be obtained but only reviewed if other areas fail to bear fruit. The relevant data includes:

- Connection history, ideally including IP addresses.
- Time of temperature change, which indicates a guaranteed timeframe the attacker was known to be in the system and is likely to be logged. This data may also be stored in the cloud, making modification far more complex.

Notably, the thermostat should be the first targeted device for recovery – this is due to the fact that if a forensic image is successfully obtained, it allows for it to be disconnected or reset with new firmware, meaning the home owner can reduce the temperature from 85C, relieving significant distress.

2.3.3 Wireless Router & Firewalls

Detailed packet capture information is unlikely to be present, as the devices are ostensibly “consumer” grade due to implementation in a smart home, though this would be ideal if it was present and would allow for analysis through programs such as Wireshark and tshark. This was assessed to be unrealistic given the scenario described.

Firewall evidence is likely to closely mirror that of router, with limited differences hence why they are counted together.

2.3.3.1 Non-Volatile Evidence

The wireless router will contain an operating system image stored in ROM; this may be proprietary or not, but can likely be obtained alongside the bootloader. In the absence of that, start-up configuration files can provide insight into potential breach vectors due to misconfiguration and will always be present, likely with them being visible through the router interface.

Firewall rules and exceptions should be examined to identify potential breach vectors as a result of improper security configuration. A firewall is also likely to retain access logs with more frequency than a router. To summarize, relevant persistent/non-volatile evidence includes:

- Operating system image
- Bootloader
- Start-up configuration files
- Access logs
- DHCP logs
- Firewall configuration

2.3.3.2 *Volatile Evidence*

Volatile evidence is more pertinent as it relates to elements that are likely to have been changed – such as the currently running config of which it is known the attacker modified to change the Network ID. The command history may provide significant information as a result of this. Furthermore, elements within this area are far harder for an attacker to theoretically modify, such as I/O and processor memory if they can be obtained.

The most significant amount of data for this investigation will be found in the Firewall and Router information due to the fact they are likely to log more information by default and the router specifically facilitates all connections to external devices to which this investigation relates. Relevant data to be recovered includes:

- Routing Tables
- Command history
- Stored packets
- Flow Data
- ARP table
- I/O and processor memory

Critically, the fact the attacker was present on the router means they may have left a trail behind, but similarly must mean that the investigator cannot reliably trust the data contained therein – as it could've been modified to cover evidence from the breach that would've been visible otherwise and it must be carefully scrutinized.

This is indicated below in the systematic diagram (Figure 4) based on (Dorai, et al., 2018).

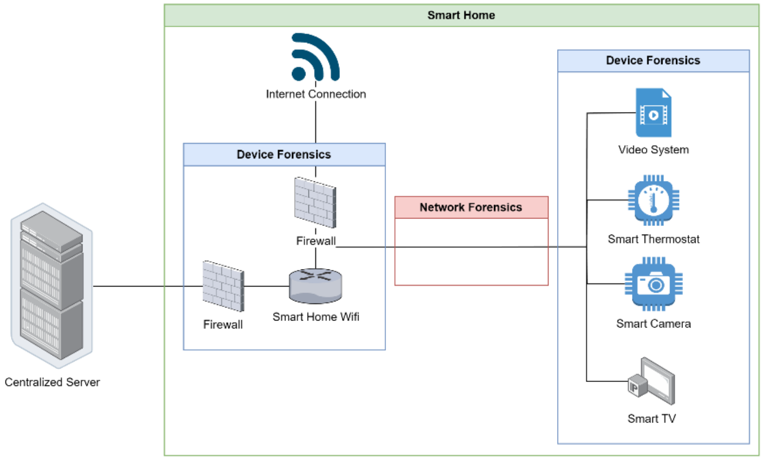


Figure 4 - Indication of IoT forensic ecosystem

2.4 ANALYSIS OF DATA

2.4.1 Local Data Analysis

Local Data Analysis is performed first here as it is of arguably lower complexity, primarily relating to the analysis of files, device logs and memory.

2.4.1.1 Memory Analysis

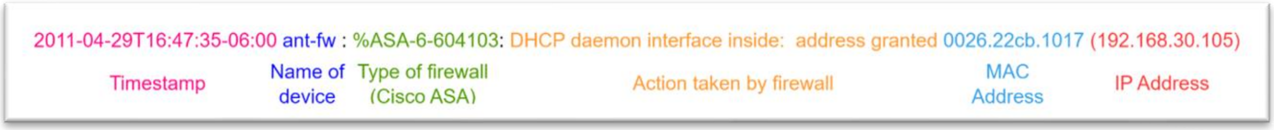
If memory data can be obtained, the tool used for analysis is typically The Volatility Framework, which depending on the underlying operating system can be employed to identify a lot of data. This can include the active network connections which could indicate a remote attacker, or network packets in kernel memory. The process tree within Linux could also indicate malware by showing running processes and files they're currently using, indicating if the attacker has attempted some form of persistence. This can be seen below in [FIGURE] where the running processes of a linux image are displayed (Sample Linux memory image obtained from the Honeynet project).

[INSERT FIGURE HERE]

2.4.1.2 Access Log Analysis

2.4.1.3 DHCP Log Analysis

DHCP refers to Dynamic Host Configuration Protocol, which is the process through which an IP address is assigned to devices on a network. This will likely contain a fairly small amount of data, but may identify an attacker by retaining the MAC and IP address that can subsequently be used to identify other areas of exploit by using it as a filter, especially if the IP address of known devices can be identified and subtracted from the log. Based on the fact that the firewalls exist on the perimeter, the attacker will have had to traverse them at some point meaning that an entry will have likely been logged. An example of how a DHCP log entry from a firewall could be deconstructed to ascertain this information can be seen below in Figure 5.



```
2011-04-29T16:47:35-06:00 ant-fw : %ASA-6-604103: DHCP daemon interface inside:  address granted 0026.22cb.1017 (192.168.30.105)
```

Figure 5 - Deconstruction of DHCP log alert from Firewall

2.4.1.4 Firewall configuration analysis

2.4.2 Network Data Analysis

2.4.2.1 Port and Protocol Analysis

The most common kind of protocol analysis refers to application layer protocols, such as HTTP, FTP and SSH. This typically maps linearly onto port designations, wherein for instance SSH is typically port 22. Based upon this, a forensic investigator can identify traffic that may be malicious in isolation, for instance, an SSH connection may be anomalous and indicate some form of external access or identify traffic making use of non-standard port/protocol designations such as a HTTP service running on any port other than 80, or an SSH protocol running on a port other than 22. One protocol of specific note is MQTT (Message Queuing Telemetry Transport), which is typically used for Machine to machine transport and is often employed with IoT devices – though this is of low exploitability.

2.4.2.2 TCP Flags

If packets are obtained, they will contain TCP flags. The three most notable TCP flags that could be useful for investigation are:

- SYN – The first packet initiating a new TCP connection, possibly indicating an initial point of compromise.
- FIN – Indicates a device has finished sending data, useful for the same reason.
- RST – Used to tear down an existing TCP connection, sent as a response in a packet if a port is closed.

This is an effective method of detecting network scanning prior to attack, as if packets are present that have a NULL flag set, or have multiple flags set such as URG and FIN simultaneously it can be indicative of NULL scanning or Xmas scanning to identify open ports – with the source of this scanning likely being the attacker and thus an effective method of identifying the source of the breach.

2.4.2.3 Statistical Flow Analysis

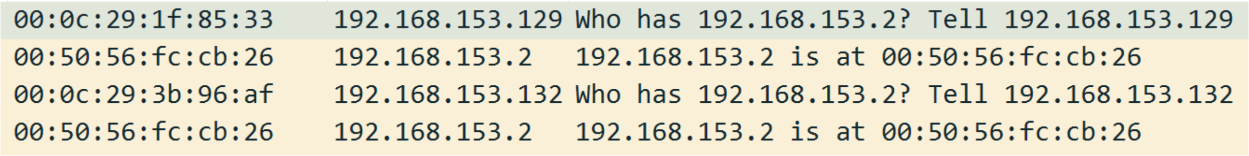
Statistical flow analysis may be theoretically possible, depending on the data retained by the router in the absence of dedicated appliances and the lack of an enterprise-grade switch. Assuming this is the case, it would be fed into the tool "YAF"(Yet Another Flowmeter) as this would allow for a number of subsequent analysis techniques to be employed using the additional tool "SILK" (System for Internet-Level Knowledge). This would pull together a lot of the prior techniques and display them in a more understandable format.

For example, initially, if a rough timeframe for the breach is possible to be obtained, historical baselining may be possible wherein "normal" network traffic is used as a baseline to identify discrepancies indicating compromise. From that, filtering can be employed to find "dirty values" – such as IP addresses, ports or protocols that indicate compromise. Most detected behaviors are variations upon this, typically searching for specific patterns within traffic based upon the array of network information which can allow for automatic identification of compromise or scanning prior to them. One of the easiest methods of doing this is to simply search for large volumes of data, often indicative of data infiltration or exfiltration – an example of this can be seen below in figure using the command:

```
rwstats capture.rw --fields=3 --values=packets --count=10
```

2.4.2.4 ARP tables

An ARP (Address routing protocol) table provides a table of all known MAC addresses and their associated IP addresses (see Figure 6). This is a low hanging piece of evidence, but can quickly provide the source of malicious data in some instances.



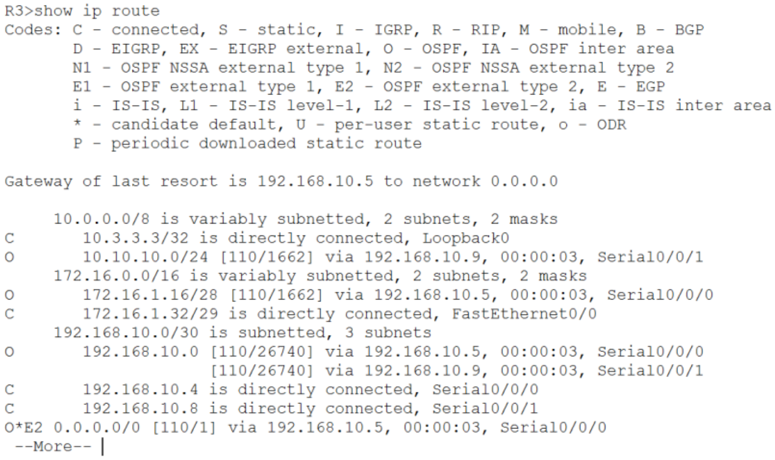
```
00:0c:29:1f:85:33    192.168.153.129 Who has 192.168.153.2? Tell 192.168.153.129
00:50:56:fc:cb:26    192.168.153.2   192.168.153.2 is at 00:50:56:fc:cb:26
00:0c:29:3b:96:af    192.168.153.132 Who has 192.168.153.2? Tell 192.168.153.132
00:50:56:fc:cb:26    192.168.153.2   192.168.153.2 is at 00:50:56:fc:cb:26
```

Figure 6 - ARP table within packet capture

One potential concern is that this is vulnerable to ARP spoofing, wherein an attacker modifies an ARP table such that their MAC address is associated with the IP of a known host, meaning all data is sent to their device. This means that the MAC address of all known devices should be ascertained locally then checked against the ARP table to identify discrepancies.

2.4.2.5 Routing tables

A routing table is used to show all of the routes to the destinations in the network. This allows for an investigator to see a malicious attacker acting as a middleman, sending packets to their intended destination – meaning behaviour may appear as normal, but in reality the data is being intercepted. This also allows for the prior method of discrepancy identification by spotting unknown IP's. An example of doing this is shown below in Figure 7 on a Cisco router with access to the CLI.



```
R3>show ip route
Codes: C - connected, S - static, I - IGRP, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2, E - EGP
       i - IS-IS, L1 - IS-IS level-1, L2 - IS-IS level-2, ia - IS-IS inter area
       * - candidate default, U - per-user static route, o - ODR
       P - periodic downloaded static route

Gateway of last resort is 192.168.10.5 to network 0.0.0.0

     10.0.0.0/8 is variably subnetted, 2 subnets, 2 masks
C       10.3.3.3/32 is directly connected, Loopback0
O       10.10.10.0/24 [110/1662] via 192.168.10.9, 00:00:03, Serial0/0/1
     172.16.0.0/16 is variably subnetted, 2 subnets, 2 masks
O       172.16.1.16/28 [110/1662] via 192.168.10.5, 00:00:03, Serial0/0/0
C       172.16.1.32/29 is directly connected, FastEthernet0/0
     192.168.10.0/30 is subnetted, 3 subnets
O       192.168.10.0 [110/26740] via 192.168.10.5, 00:00:03, Serial0/0/0
                     [110/26740] via 192.168.10.9, 00:00:03, Serial0/0/1
C       192.168.10.4 is directly connected, Serial0/0/0
C       192.168.10.8 is directly connected, Serial0/0/1
O*E2 0.0.0.0/0 [110/1] via 192.168.10.5, 00:00:03, Serial0/0/0
 --More-- |
```

Figure 7 - Show IP Route on a router

3 DISCUSSION

3.1 CRITICAL EVALUATION

Data acquisition issues related to proprietary technology have already been mentioned.

One element of improvement in relation to the topology is that based on the description it seems likely the firewall is external facing, with the router – this is fine given that

But it means that if the router is attacked, the firewall

It also increases the risk of denial of service against a firewall,

3.2 REFLECTION

Aside from those previously mentioned, other methods of obfuscation performed by an attacker can impede the investigation. Potentially relevant elements to this investigation are:

- Accessing the router allowing for local data modification, log deletion etc – general interference with standard behavior as a result of obtaining access during a breach
- Long standing access, meaning that the breach could have occurred a long time ago and only been utilized recently – meaning logs may not be kept that far back by default and "baselining" may not be useful

4 CONCLUSION

In conclusion, due to the nature of the evidence likely to be realistically obtainable in the network infrastructure a successful digital forensic investigation can likely be undertaken. There are some significant challenges such as the contamination of arguably the greatest data source (the router used as the central communication point between IOT devices) and the devices in question potentially preventing the obtainment of forensic images due to their proprietary nature. However, while no strict packet capture data is likely to be retained owing to the consumer nature of the devices, the retained information in the form of DHCP information, stored packets and router modification history among other information should allow for the investigator to ascertain the source and nature of the breach.

5 References

DOMO, 2018. *Data Never Sleeps 6.0.* [Online]
Available at: https://www.domo.com/assets/downloads/18_domo_data-never-sleeps-6+verticals.pdf
[Accessed 15 December 2024].

Dorai, G., Houshmand, S. & Baggili, I., 2018. *I Know What You Did Last Summer: Your Smart Home Internet of Things and Your iPhone Forensically Ratting You Out.* Hamburg, Association for Computing Machinery, pp. 1 - 10.

Findlay, B., 2021. A forensically-sound methodology for advanced data acquisition from embedded devices at-scene. *Forensic Science International: Reports,* Volume 3.

Gandhi, V., 2023. *2023 ThreatLabz Report Indicates 400% Growth in IoT Malware Attacks.* [Online]
Available at: https://www.zscaler.com/blogs/security-research/2023-threatlabz-report-indicates-400-growth-iot-malware-attacks
[Accessed 16 December 2024].

Sinha, S., 2024. *State of IoT 2024: Number of connected IoT devices growing 13% to 18.8 billion globally.* [Online]
Available at: https://iot-analytics.com/number-connected-iot-devices/
[Accessed 16 December 2024].

Zawoad, S., Hasan, R. & Skjellum, A., 2015. *OCF: An Open Cloud Forensics Model for Reliable Digital Forensics.* New York, Institute of Electrical and Electronics Engineers.

APPENDICES

APPENDIX A