Grey Areas of Digital Forensic Tools and a Framework to Solve the IoT Data Forensic Analysis Problems

Benjamin Benson[1] and Gokila Dorai[2]

[1] Augusta University, Augusta, Georgia
bebenson@augusta.edu
[2] Augusta University, Augusta, Georgia
gdorai@augusta.edu

Abstract

The constant and rapid evolution of technology has led to some amazing achievements. Normal people can communicate with others across the globe, relatively cheap Internet of Things (IoT) devices can be used to secure homes, track fitness and health, control appliances, etc., many people have access to a seemingly endless wealth of information in small devices in their pockets, organizations can provide high availability for important services by spinning up/down servers in minutes to scale with demand through cloud services, etc. However, not everyone who uses these technologies does so with a pure heart and good intentions, many people use them to commit or help commit crimes. A nefarious individual may use cloud services to host a highly available Command and Control (C2) server, a messaging app to form and communicate with a gang or hacking group, or IoT devices as part of a botnet designed to perform Distributed Denial of Service (DDoS) attacks. When these technologies are used in the commission of a crime, they hold valuable information that needs to be recovered forensically to use as evidence to convict the perpetrators. Unfortunately, that ever-evolving technology poses many challenges for digital forensics. This paper identifies and presents many of the challenges faced in digital forensics involving mobile devices, IoT devices, and cloud services in addition to proposing a framework for solving the IoT Forensic Data Analysis problem.

Contents

1 Microservices Platforms and Digital Forensics Challenges

Containerization with microservices solves the challenge of delivering a good end-user experience by allowing IoT devices to group the resources to be shared as and when needed. However, IoT devices interact heavily with companion/controlling devices through mobile applications, web applications and cloud components. In the following subsections, we explore the digital forensics challenges associated with mobile devices, IoT devices, and cloud applications that could impact data forensics and analysis of IoT data. In our previous works, we have identified issues related to IoT forensics [3], the enormous amount of user data that can be recovered from

IoT applications using digital forensic acquisition and analysis methods [4] and an overview of targeted data extraction system using applied machine learning models depending on the contents of mobile applications [2]. Finally, in section 2, we propose a framework to address these challenges by using sensor data classification and analysis on edge devices using machine learning combined with sensor semantic annotations.

1.1 Mobile Devices and Digital Forensics Challenges

The forensic investigations of mobile devices by law enforcement agencies usually involves the use of commercial software and/or hardware tools that extracts and analyzes data from the device, but utilizing these tools is not always straightforward and easy. One of the biggest roadblocks to extracting and analyzing the data on these devices are that their software (the mobile operating system and the many installed apps) are constantly being updated [5]. These updates can easily change what data is collected and stored on the device, the format of the data, the logical location of the data in the file system, or the interface that allows access to the data. Any or all those changes could prevent the tools used by law enforcement, such as Cellebrite, from extracting or analyzing all the relevant evidence from the device [6]. If this happens, they will have to rely on someone with the skills and knowledge to manually extract the evidence or wait for the vendor to update the tool. Another issue they may face is that many of these tools must exploit a few vulnerabilities to access the data on the device if the digital investigator does not have the credential to access the device. The vulnerabilities they use could be patched in the next update which would mean that the digital investigator must wait for the tool to be updated with new exploits. Furthermore, there are far too many apps, that may harbor useful information, for an investigator or vendor to be sure that they have the knowledge or tools to extract evidence from every app that may be on a device.

1.2 IoT Devices and Digital Forensics Challenges

The moniker IoT encompasses a plethora of very diverse devices, including smartwatches, smart thermostats, security cameras, motion sensors, smart lights, smart appliances, fitness trackers, GPS tracking devices, doorbells, speakers, etc. The protocols used by these devices are diverse in nature and sometimes very specific to the vendor [1]. This diversity is one of the challenges investigators face when trying to gather evidence. It is nearly impossible for a digital investigator or tool to be able to identify every possible IoT device and extract useful information from it. Each device can use a different networking protocol, collect different data, store data differently, collect a ton of data to stream or very little sensor data. Also, these devices do not usually have much on board storage, and instead rely on cloud storage or other mobile devices. This also poses a challenge as the investigator now has to figure out where each of the devices stores data in the cloud and try to access it with warrants/subpoenas or get the credentials from the owner if possible. IoT is one of the newer fields in technology and as such it is evolving and changing more drastically than other spaces, making it harder for digital investigators and commercial forensic tool vendors to keep up with the fast evolving IoT/IoE technological advancements.

1.3 Digital Forensics Challenges Associated with Cloud Apps and Services

Cloud services are another rapidly emerging technology that imposes many challenges for digital forensics investigators. Cloud services are usually multi-tenant, meaning that their servers and

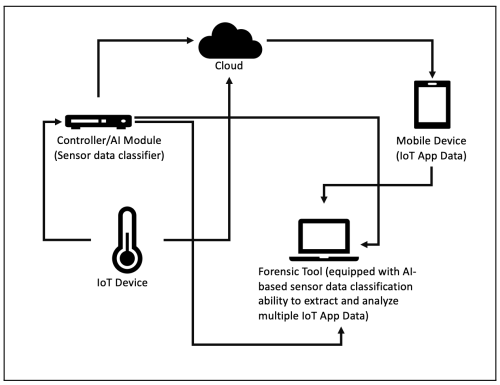

Figure 1: Overview of the Proposed Framework

infrastructure serves many different customers. One challenge this can cause is that it can be hard to investigate without breaching an unrelated tenant's confidentiality, as investigators may need to physically acquire data storage devices that have data from many tenants that they are not investigating. Criminals may also escape the virtualization utilized by cloud providers and access other tenants' resources making it look like they performed the criminal activity. Collecting physical evidence from cloud providers can be difficult because the evidence may be far away, possibly outside of the investigator's jurisdiction requiring them to rely on the cloud provider or other law enforcement agencies to give them access or acquire the evidence for them. Criminals can also increase the surface area that needs to be investigated by utilizing different cloud providers for different aspects of their criminal activity (i.e., AWS for the C2 server and Azure for the database and storage), slowing down investigations.

However, with many cloud services, the user's devices/clients have access to much of the relevant evidence. For example, with most smart thermostats the temperature and usage history are stored in the cloud but viewable on the user's device. By targeting the cloud service's user's devices and doing more thorough forensics on them, the many hurdles of cloud forensics can be cleared easily. The digital investigator does not have to worry about infringing on the privacy of another tenant or getting physical drives from a cloud provider.

2 Proposed Solution

To help digital investigators extract and analyze the diverse and useful information on mobile devices, we propose the use of an edge-based machine learning software tool that can extract and interpret IoT data from mobile devices (including data stored in the cloud that is accessible through the mobile device) and present the data in a useful way. The software tool would be deployable on devices such as a laptop or Raspberry PI. It would be trained to classify data that it extracts from mobile devices so the digital investigator can select which types of data they need. For the IoT and cloud data, it would retrieve what it stored on the device and attempt to gain access to the remotely stored data by retrieving authentication tokens. It will use machine learning to attempt to classify what kind of sensor(s) the IoT data came from and organize it in a useful way. Once it collects and classifies the data it will present it in a useful format, such as a graph for temperature, a map for geolocation data, a table timeline with datetime for a motion sensor, a timeline with all or most of the data, etc.

Congruently, the framework will be equipped with an IoT traffic controller software tool

that will intercept and collect the IoT device's network communications before they are sent to the cloud. This will also require hardware that can interact with the many different networking protocols used by IoT devices, such as ZigBee, z-wave, Weave, etc. This component would just collect the data and provide an easy way for digital investigators and/or the other component of the tool to extract useful information from IoT devices without the need to process data from mobile devices or cloud storage.

To summarize, classical techniques used in information retrieval for searching and retrieving information are not helpful in certain scenarios because of the heterogeneous and voluminous nature of sensor data. Using enhanced descriptions of sensors, a set of measurement data, and a record of observations, combined with sensor semantic annotations, our framework will classify sensor data and provide meaningful insights overlayed on a timeline view.

References

[1] About z-wave technology. https://z-wavealliance.org/about_z-wave_technology/. Accessed: 2022-03-04.

[2] Sudhir Aggarwal, Gokila Dorai, Umit Karabiyik, Tathagata Mukherjee, Nicholas Guerra, Manuel Hernandez, James Parsons, Khushboo Rathi, Hongmei Chi, Temilola Aderibigbe, et al. A targeted data extraction system for mobile devices. In *IFIP International Conference on Digital Forensics*, pages 73–100. Springer, 2019.

[3] Gokila Dorai, Shiva Houshmand, and Sudhir Aggarwal. Data extraction and forensic analysis for smartphone paired wearables and iot devices. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*, 2020.

[4] Gokila Dorai, Shiva Houshmand, and Ibrahim Baggili. I know what you did last summer: Your smart home internet of things and your iphone forensically ratting you out. In *Proceedings of the 13th International Conference on Availability, Reliability and Security*, pages 1–10, 2018.

[5] Samiha S Shimmi, Gokila Dorai, Umit Karabiyik, and Sudhir Aggarwal. Analysis of ios sqlite schema evolution for updating forensic data extraction tools. In *2020 8th International Symposium on Digital Forensics and Security (ISDFS)*, pages 1–7. IEEE, 2020.

[6] Ariel Watson. Ok google is more than ok for digital intelligence investigations. https://cellebrite.com/en/ok-google-is-more-than-ok-for-digital-forensics-investigations/, 2018. Accessed: 2022-03-04.