Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Dept. of Information Security and Communication
Technology

Daniel Bing Andersen

Tool induced biases in iOS pattern of life analysis

Master's thesis in MISEB
Supervisor: Kyle Porter
Co-supervisor: Nina Sunde

December 2023

NTNU
Norwegian University of
Science and Technology

Daniel Bing Andersen

Tool induced biases in iOS pattern of life analysis

Master’s thesis in MISEB
Supervisor: Kyle Porter
Co-supervisor: Nina Sunde
December 2023

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Dept. of Information Security and Communication Technology

NTNU
Norwegian University of
Science and Technology

Abstract

Mobile devices and smart watches store increasingly more information about users' activities, capturing everything from location and health data to connection and device usage history. In digital forensic circles, the process of analyzing such data has coined the term *pattern of life analysis*, suggesting a richness of data capable of reproducing aspects of a user's digital, physical and even mental activities.

Among other things, law enforcement takes advantage of this granularity of data to verify information provided in interviews, to establish sequences of events, and to reveal potential intents. Digital forensic investigators mainly rely on proprietary analysis tools for such inquires. However, misinterpretation of evidence and access to partial information can lead to miscarriages of justice.

On this bases, the study seeks to answer whether the presentation layer of digital forensic tools can lead to errors and misinterpretations. Through a comparative analysis of the digital forensic analysis tools Cellebrite Physical Analyzer and APOLLO, two groups of digital forensic investigators performed a series of assignments, each group with results from one of the two tools.

The study found a notable discrepancy between the performance of the groups. Errors in the assessments were classified based on factors of influence within the tools' presentation layers, including naming conventions, details, data aggregation, and timestamps.

Earlier research have shown how biases can affect forensic experts in their opinion making processes. However, the study found a source of bias that has not previously been examined within the context of digital forensics. Therefore, a classification of *data presentation biases in forensic analysis* is introduced as a way to concretize these findings.

Sammendrag

Mobile enheter og smartklokker lagrer stadig mer informasjon om brukerens aktiviteter. Dette kan inkludere detaljer rundt stedstjenester, tilkoblinger, helsedata og enhetsbruk. Analysen av slike bruksdata har fått betegnelsen *pattern of life analysis* i det digitale rettstekniske miljøet, et begrep som peker mot en detaljrikdom som er tilstrekkelig for å gjenskape aspekter av brukerens digitale, fysiske, og mentale aktiviteter.

Dataetterforskere har blant annet benyttet slike data til å verifisere informasjon som fremkommer under avhør, for å gjenskape hendelsesrekker, og for å påvise forsett. Dataetterforskere benytter i all hovedsak proprietære analyseverktøy til slike formål. Feiltolkninger og tilgang til ufullstendige data kan imidlertid føre til justisfeil.

Studien søker derfor å svare på om måten analyseverktøy fremstiller data på kan føre til feiltolkninger. For å svare på dette ble det gjennomført en komparativ analyse av to digitale etterforskningsverktø Cellebrite Physical Analyzer og APOLLO. To grupper med dataetterforskere ble hver gitt resultater fra et av de to verktøyene, og ble deretter bedt om å besvare en serie med informasjonsbehov.

Studien fant betydelige ulikheter mellom gruppenes evne til å svare på informasjonsbehovene. Feil og mangler som oppstod undergjennomføringene ble klassifisert ut fra hvilken type påvirkning verktøyenes presentasjonslag hadde på resultatene. Klassifiseringen inkluderer navnekonvensjoner, detaljer, aggregering av data, og tidsstempler.

Eksisterende forskning viser hvordan eksperter kan påvirkes av bias i beslutningsprosesser. Studien har imidlertid avdekket en kilde til bias som ikke har blitt undersøkt tidligere i digital rettsteknisk kontekst. For å konkretisere funnene defineres klassifiseringen av denne feilkilden som *data presentation biases in forensic analysis.*

Acknowledgments

I want to express my gratitude to all the individuals who contributed in the making of this thesis:

First of all I want to thank my supervisor from NTNU, Kyle Porter, for his guidance and support.

A special thanks goes to my co-supervisor Nina Sunde at Politihøgskolen, who contributed with invaluable support in finalizing the thesis. I also want to acknowledge her research, as it was an inspiration when selecting the theme of the thesis.

A big thank you to all of the participants who dedicated time and efforts to the experiment, your contributions serve as the backbone of this study.

Finally, I need to thank my sister Solvår who facilitated the arrangements for the scenario, and who helped me to commit a ”murder” for the sake of science.

Table of Contents

Abstract 1

Acknowledgments 3

List of Figures 7

List of Tables 7

1 Introduction 1

1.1 Motivation 1

1.2 Research problem 2

1.3 Contributions 2

1.4 Thesis structure 3

2 Background and related work 4

2.1 Digital evidence and policing 4

2.2 Digital forensics and the forensic sciences 5

2.3 How digital forensic opinions are made 6

2.4 Pattern of life analysis 6

2.5 Data at the activity data 7

2.6 Pattern of life analysis in criminal investigations 8

2.7 Related work 9

3 Method 11

3.1 Research methodology and philosophy 11

3.2 Research design 11

3.3 Delimitations 12

3.4 Sampling 13

3.5 Tool selection 13

3.6 Hardware 14

3.7 Pilot study 14

3.8 Data creation 15

3.9 Data extraction 15

3.10 Data reduction 20

3.11 Data processing and normalization 21

3.12 Data analysis 24

3.13 Biases . . . 25

3.14 Strengths and weaknesses . . . 26

4 Experiment 27

4.1 Data Creation . . . 27

4.1.1 External devices . . . 27

4.1.2 How the scenario was created . . . 29

4.1.3 Detailed timeline . . . 34

4.2 Background information given to the respondents . . . 35

4.3 Information needs . . . 35

4.4 Interview . . . 36

5 Results and analysis 38

5.1 Participants background and experience . . . 38

5.1.1 Background and education . . . 38

5.1.2 Familiarity with forensic tools . . . 38

5.1.3 Experience with data at activity level . . . 39

5.2 Interpretation of results . . . 39

5.2.1 Task 1: Communications . . . 40

5.2.2 Task 2: Connections to Wi-Fi networks . . . 41

5.2.3 Task 3: Bluetooth connections . . . 42

5.2.4 Task 4: Car Play . . . 45

5.2.5 Task 5: Browser history . . . 46

5.2.6 Task 6: Original of photo . . . 47

5.2.7 Task 7: Connection to charger . . . 49

5.2.8 Task 8: Last activities . . . 50

5.2.9 Task 9: Airplane mode . . . 52

5.3 Post-scenario interviews . . . 52

5.3.1 Technical challenges . . . 52

5.3.2 Missing data . . . 53

5.3.3 Interpretation . . . 53

5.3.4 Accuracy and reliability . . . 54

5.3.5 Tool selection . . . 54

5.3.6 Competency . . . 55

5.3.7 Additional comments . . . 56

6 Discussion **56**

6.1 Factors of the tools interpretation layers 56

6.1.1 Naming conventions 56

6.1.2 Details and contextual information 58

6.1.3 Data aggregation 61

6.1.4 Timestamps 64

6.2 Group performance 66

6.2.1 Where did the groups perform on par? 66

6.2.2 The essence of differences 66

6.2.3 Measurement of performance 67

6.2.4 Conclusion of the performance measure 69

7 Conclusion **71**

References **73**

Appendices **75**

1 Sikt approval 76

2 Consent form 78

3 Question form 80

4 Interview guide 82

5 Activity types APOLLO 83

List of Figures

1 Unc0ver jailbreak loaded on device 16

2 Unc0ver jailbreak process completed 16

3 Device successfully jailbroken with Unc0ver 17

4 iProxy started on port 22 44 18

5 SSH root access to the device 18

6 SSH root access to the device, list list of files 18

7 iOS Triage full filesystem extraction started 19

8 iOS Triage full filesystem extraction completed 19

9 XML structure for modifying properties in APOLLO 23

10 Setup of Wi-Fi network named Berit Knausen WIFI 28

11 Setup of Bluetooth speaker named JBL Flip 5 28

12 Setup of Apple CarPlay, named Skoda EL25467 29

13 Scenario, at parking lot outside health center 12:24 30

14 Scenario, entering restaurant at 13:15 30

15 Scenario, eating at restaurant 13:40 31

16 Scenario, leaving restaurant at 14:41 31

17 Scenario, playing music with Bluetooth speaker at 15:57 32

18 Scenario, photo 1 taken with the device at 18:07 32

19 Scenario, photo 2 taken with the device at 19:53 33

20 Scenario, photo of the device left on the ground, taken at 19:54 33

21 Distribution of errors in the Cellebrite group 68

22 Distribution of instances in the APOLLO group 69

List of Tables

1 Databases and number of rows processed by Cellebrite and APOLLO 20

2 Excerpt of top 25 activity types in APOLLO 21

3 Activity types in Cellebrite 21

4 Excel spreadsheet exported from Cellebrite's timeline view 22

5 Excel spreadsheet from Cellebrite timeline, normalized to its final form 22

6 APOLLO encloses attributes within square brackets 23

7 APOLLO normalized output 24

8 APOLLO normalized output in its final form 24

9 Timeline for scenario 35

10 Detail level of activities in Cellebrite 44
11 Detail level of activities in APOLLO 44
12 Detail level of web history in Cellebrite 47
13 Detail level of web history in APOLLO 47
14 Activity sensor data in Cellebrite 51
15 Device events in Cellebrite 57
16 Device Events in APOLLO 58
17 Harvested WiFi locations in Cellebrite 60
18 Harvested WiFi locations (without coordinates) in APOLLO 61
19 Camera usage in Cellebrite 64
20 Camera usage in APOLLO 65
21 Aggregated health data in Cellebrite 65
22 Synthesized elements in the tools presentation layers 67
23 Errors encountered in the Cellebrite group 68
24 Errors encountered in the APOLLO group 69
25 Activity types and rows for APOLLO 83

1 Introduction

Evidence seized from mobile devices plays an important role in today's criminal investigations. With time, mobile phones have transformed into handheld logging devices containing vast amounts of information about movement, network connectivity, usage patterns, and location history, data that could play an incriminating role in a subsequent forensic analysis.

In digital forensic circles, the review of such data is often referred to as pattern of life analysis. The terminology, which originated in intelligence communities (Craddock et al., 2016), is used to describe the fact that the amount of available data can not only give detailed insight of the device usage, but to the degree by which the user is tied to their devices, a chronological insight of the users digital, physical, and even mental behavior. In Norway alone, the theme of device usage, health data and location history has been a subject matter in several criminal cases (TOSLO-2020-20518-2; TOSL-2022-14854; LB-2022-17642).

An accurate analysis of device usage is a time-consuming endeavor that requires highly specialized skills. Therefore, digital forensic investigators mainly rely on semi-automated proprietary tools for their analysis. In recent years, viable open source options for pattern of life analysis have emerged, producing detail levels that are comparable to, or even surpassing that of the market-leading proprietary tools.

As pattern of life analysis is increasingly submitted as evidence in court, questions regarding accuracy of data, interpretation of artifacts, and transparency by which the results were obtained should be raised. As often is the case, actors in the court have limited experience with interpretation of digital traces (Erlandsen, 2019). As such, digital forensic investigators bear a significant responsibility to conduct their investigations in a way that safeguards the rule of law.

Previous research has found inconsistencies in how digital investigators interpret data (Sunde, 2021; Sunde and Itiel E. Dror, 2021). As pattern of life analysis typically involve an analysis of contextual information to highlight sequences of events, results are often used to support overriding evidence (Meconi and Henseler, 2022). Due to the interpretative narratives that could occur, it is of concern that it is particularly susceptible to human and tool induced errors.

On this basis, the thesis seeks to investigate the intersection between human and tool induced biases, specifically regarding the way in which forensic tools presents their results. By conducting a comparative analysis between the market leading mobile forensic analysis tool Cellebrite Physical Analyzer (Cellebrite, 2023) and the open-source tool APOLLO (Edwards, 2020a), the aim is to identify whether inconsistencies in results can be attributed to elements within the tools' presentation layers.

1.1 Motivation

Digital forensic investigators are commonly required to present their findings in court. With specialization in the field, they are expected to provide insight into the inner workings, intended purpose, reliability, and provide opinions about the evidential value of particular artifacts. In addition, investigators should comply with principles of accountability (Williams, 2012), as results should be reproducible with methods open for scrutiny, a paradox seen in the light of proprietary tools.

As cases treated in the legal system have real impacts on human lives, missing data or wrongfully interpreted evidence could lead to miscarriages of justice. In Norway alone several cases of wrongful imprisonments, and years of public accusation are haunting the legal system. The concern is that the increased reliance of pattern of life analysis in court could if not treated soundly, add to this series of miscarriages.

With support for all major mobile operating systems, Cellebrite stands as the de facto forensic analysis tool within law enforcement agencies. Due to its support for a wide array of devices, the tool implements strict data normalization. This data normalization process entails a confinement of parsed data into a limited set of categories. As a result, unified categories contain data with varying degrees of reliability, with few distinctions other than their source file reference. An additional effect of the tool's normalization process, is that data are parsed into a restricted set of attributes, offering a simplified view of the source data. Although similar proprietary tools share the same characteristics, the tool is of special interest due to its widespread recognition within law enforcement.

The claims mentioned above do not appear from a vacuum, but are based on my day to day experience as a digital forensic practitioner. They are also supported by my own experiences of peer reviewing reports, paired with feedback and discussion with multiple fellow practitioners.

In recent years, several open source alternatives have emerged. My personal belief, supported by Carrier, 2002, is that open source tools better adhere to basic principles of digital forensics, such as openness, transparency and accountability of results.

A viable open source alternative for pattern of life analysis is APOLLO, a Python script developed by Sarah Edwards (Edwards, 2020a). The tool poses radically different characteristics than Cellebrite, as there are no inherent limitations regarding granularity of activities and attributes. With a granularity of artifacts that is comparable to, or superseding that of Cellebrite, it is of interest to the research problem stated in next section.

As detailed in the above, the motivation behind the thesis stems from an overarching concern regarding the way in which proprietary forensic tools present their data. The aim is to shed light on potential factors of misinterpretation, and to provide a framework that could help mitigate errors in the future.

1.2 Research problem

The use of pattern of life analysis as evidence in court has become a common practice. Given that such analyses are heavily dependent on contextual information, there is a concern that it is particularly vulnerable to misinterpretations. Since digital forensic investigators primarily rely on proprietary semi-automated tools, it is of interest to explore in what way their presentation of data affect investigator opinions.

In order to answer the research problem, the following research questions have been formulated:

1. What factors within the presentation layer of digital forensic analysis tools affect investigator opinions?
2. If such factors are found, to what extent do they affect criminal investigations?

1.3 Contributions

In contrast to sources of bias covered by earlier studies (Cooper and Meterko, 2019; I. E. Dror and Mnookin, 2010; Itiel E Dror, 2020; Itiel E. Dror et al., 2012; Erlandsen, 2019; Smit et al., 2018), the study focuses on a novel area of biases that lie in the intersection between human and tool-based errors, more specifically, how the presentation layer in digital forensic analysis tools affects experts in their opinion making.

The thesis contributes a concrete classification of such factors, termed *data presentation biases in forensic analysis.* This classification can be expanded and refined in future studies, and could contribute to design choices in the future. With a concrete classification, the digital forensic community given a tool whereby this category of biases can be identified, contributing to awareness and increased quality in digital forensic decision making processes.

1.4 Thesis structure

The chapters of the thesis are organized in the following manner:

Background and related work:
The reader is given an insight into the current state of digital forensics, its scientific standing and how the concept of pattern of life analysis has been incorporated within the field. It examples the use of pattern of life analysis in real-world cases, and provides related research.

Methodology:
The chapter provides an overview of the study's methodology and research design. It accounts for hardware and software choices, and details the data acquisition, processing and normalization phase.

Experiment:
Due to its comprehensive nature, details regarding the data creation and collection phase are placed in its own chapter. The chapter accounts for the creation of the study's ground truth and details participant assignments and the post-scenario interviews.

Results and analysis:
Data gathered through participant assignments and interviews are presented and analyzed. Each assignment is analyzed independently to help identify factors of influence within the tools presentation layers.

Discussion:
The chapter includes a synthesis and discussion of factors found throughout the previous chapter. A performance measure is presented to showcase the distribution of errors between the groups.

Conclusion:
The chapter contains an overall conclusion, and links the research questions to the study's findings.

2 Background and related work

The thesis seeks to determine how the presentation layer of digital forensic tools influence investigator opinions when conducting pattern of life analysis. To address the context and relevancy of the research question, some background information is needed.

The background chapter is introduced with a brief discussion on the evolution of crime and its digital components. It showcases how these components led to the emergence of digital forensics, and that the field still strives to align itself with the scientific rigor of long-standing forensic practices. It continues to explore the area of digital forensics, and the principles by which investigators form opinions. Potential challenges of forensic decision-making processes are highlighted, before introducing the concept of pattern of life analysis. Current research in the area of pattern of life analysis is explored, before exemplifying its use cases in real-life criminal investigations. The chapter continues with a discussion on narratives and their incorporation into analysis at the activity level. The chapter concludes with related works to bring further context to the core of the research problem.

2.1 Digital evidence and policing

A majority of today's criminal investigations encompass some form of digital evidence. The National Police Chiefs Council (NPCC, 2020) has recognized that over 90% of all criminal activities have digital components. This observation is supported by Wilson-Kovacs et al. (2023), who underscores that digital evidence has surpassed other types of forensic evidence. During interviews with criminal defense lawyers, mobile devices were highlighted as the most common source of evidence, used by prosecutors in more than 50% of their cases. Despite that the use of digital evidence has been tightly integrated with law enforcement agencies, NPCC (2020) suggests that it has yet to reach its potential in commonplace technologies.

The increased reliance on digital evidence by law enforcement could be explained in a twofold facet: a broad and general emergence of technology in society at large, and the adoption of the same technologies by criminal actors.

McQuade (2006) identifies ten technology-enabled areas of illicit activities, including communications, planning and surveillance. McQuade argues that a perpetual technological competition has risen between criminal actors and law enforcement. The operationalization of emerging technologies by illicit actors thus leads to temporary periods of unbalance, where law enforcement lacks appropriate countermeasures. As a response, law enforcement evolves, and creates a temporary state of equilibrium as it fills the gaps with competing strategies. Over time, this continual cycle of technological competition leads to an environment where both crime and policing become more complex.

The phenomenon described by McQuade is not exclusive between criminals and law enforcement: A similar cycle can be observed in the relationship between law enforcement and private companies, and the former's ability to gain access to data at need. On one hand, private companies develop techniques to safeguard their customer data, on the other hand, security researchers work relentlessly to bypass the same measures. In the area of smartphone vendors, this phenomenon has resulted in a wide array of data extraction methods (Fukami et al., 2021), with most resulting from attempts to circumvent such security measures. Combined with frequent changes in both software and hardware, digital forensic investigators find themselves in a dynamic landscape that requires continual adaptation.

In context of the above mentioned landscape, it is of interest to delve into the scientific basis of field, and how it stands in comparison to other long standing forensic sciences.

2.2 Digital forensics and the forensic sciences

Digital forensics began its integration with law enforcement in the 1980s, in response to a surge in computer crimes. This led to the establishment of several national groups specializing in computer crime investigations. Due to a continual demand for such specialization, local departments were eventually established, and set the scene for the broader integration of digital forensics in law enforcement. Despite that the field had become more integrated, it was not until the 1990s that digital forensics was recognized as a separate forensic discipline (Casey, 2004).

One of the initial conceptual frameworks for digital forensics was developed by Pollitt (1995). The framework bore significance, as it was set to be both scientifically sound and legally acceptable. The framework consisted of four stages: acquisition, identification, evaluation, and admission, stages that are still recognized today. Since then, more than 32 additional frameworks have been proposed, reflecting an advancement towards greater scientific rigor in the field of digital forensics (Yeboah-Ofori et al., 2019).

Despite the field's significant advances towards a scientific foundation, its frameworks and methodologies are still new compared to other well established forensic sciences. Even though there have been efforts to integrate standardized principles on the field, it still suffers from challenges due to tool limitations and human misinterpretations. The rationale behind these challenges is arguably an insufficient consideration for the scientific method, and a tendency to treat digital evidence as factual, without sufficient scrutiny. A proposed antidote includes training that facilitates a greater understanding of the distinction between technical, investigative and evaluative considerations. Especially relevant for the thesis are the evaluative aspects, as they involves the ability to critically assess the evidence on hand by evaluating its relevance, reliability and integrity, all within a broad investigative context (Casey, 2019).

In a study conducted by Sunde and Itiel E. Dror (2021), the consistency of digital forensic interpretations were examined. 53 digital forensic practitioners were presented with the same digital evidence and asked to interpret it. Although the respondents had access to the same evidence, they were placed in groups based on their access to contextual information. Three groups were given different information, suggesting either guilt, weak guilt or innocence, while the last group was denied any contextual information as they acted as a control group. The study demonstrated significant inconsistencies between examiners retrieving different contextual information. The group that was given information suggesting guilt, reported on more traces compared to the group that received information suggesting innocence, or no information at all. What might be the study's most discerning finding, was a low consistency between examiners who had access to the same contextual information, a finding that challenges the perception of digital evidence as objective. This is especially concerning, as digital evidence often holds a status as credible and value neutral, despite that a plethora of human and technical errors could arise (Sunde, 2022).

Casey (2002) draws attention to a variety of sources that could lead to wrongful conclusions. These sources are not limited to the interpretative phase, but encompasses all stages of the investigative process: from the concealment or fabrication of evidence, to incomplete data acquisitions, erroneous logging, interpretation of evidence, and miscommunication of results. Casey emphasized that an understanding of the broad array of potential sources of errors should be integral to forensic examiners. Not only this, but it should be encompassed within a scientific basis, through systematic identification and hypothesis testing.

While the field of digital forensics is vulnerable to a broad array of potential errors, it is not unique among the forensic sciences. In a systematic review by Smit et al. (2018), 235 instances of misleading evidence were analyzed. While witness evidence accounted for 39 % of the false conclusion, forensic evidence from various disciplines followed in second place, accounting for 32 % of the instances. Based on the type of hypothesis the evidence was meant to address, 66 % of the hypothesis addressed evidence at activity level, such as violence and abuse.

Similar findings were reported in a systematic review by Cooper and Meterko (2019). The review encompassed 29 primary research studies that focuses on biases within various forensic disciplines, such as DNA, bloodstain and handwriting analysis. One of the key findings revolved around cases where forensic analyst had access to case specific information, either about the suspect or crime scenario. As was found, access to existing case-related knowledge made practitioner more vulnerable to various forms of confirmation bias.

As seen in studies by Smit et al. (2018) and Cooper and Meterko (2019), the potential for biases is not unique to digital forensics, but encompasses several forensic disciplines including document analysis, DNA, drug analysis etc. The findings of these studies thus suggests a potential for biases in all methods that rely on subjective aspects of human interpretation.

The section has briefly explored the origin of digital forensics and how it gradually was incorporated as an independent forensic discipline. It has touched upon its scientific standing, and how it still struggles to reach the scientific rigor of more long-standing forensic disciplines. Several areas for potential misconceptions were highlighted, showcasing how factors such as confirmation bias is shared among several areas that rely on human interpretation. The next section continues by exploring digital forensics, and some of the decision-making processes that are used by investigators.

2.3 How digital forensic opinions are made

In a comparative study by Sunde (2021), digital forensic reporting practices were examined. The study was based on reports by Sunde and Itiel E. Dror (2021), where participants were tasked to analyze the same evidence. There was found notable differences in how digital forensic investigators expresses uncertainty in reports. These differences were manifested by two main conclusion types, namely categorical and strength of support conclusions. While categorical conclusions provide a definitive level of certainty, strength of support conclusions include a certainty-degrees for given hypotheses. The distribution was found to be even between the two conclusion types, illustrating an ambiguity within digital forensic reporting practices. The study suggests that conclusions should be clearly articulated and provide background information on both their certainty levels and limitations. Similar conclusions are drawn by Casey (2002), who emphasizes that conclusions should be supported by a framework that takes the reliability and confidence of the findings into account.

The lack of unity in reporting practices is somewhat comparable to findings by Sunde and Itiel E. Dror (2021), where it was found that investigators provided different conclusions on the same evidence. The suggested differences in both conclusions and reporting practices underlie claims by Casey (2019), arguing that there is a lack of basic scientific grounding and established framework in digital forensics decision-making processes.

A framework of interest, that has the potential to enhanced quality in both decision-making processes and reporting, is suggested by Horsman (2022). The paper introduces the concept of scaffolding as a form of quality assurance in forensics opinion making. The framework proposes the use of three primary sources of knowledge: past experience, target testing, and the reference of existing bodies of knowledge. Importantly, assumed or untested knowledge is excluded from the process. The framework thus addresses the some of the concerns expressed by Casey (2019), as it is suggested that examiners should be able to demonstrate the scaffolding process as a form of quality control.

2.4 Pattern of life analysis

The term pattern of life analysis originates within the intelligence community, where it entails a systematic gathering of data to determine patterns of entities. Although no formal definition has been made, the intelligence community typically employs the term in a human centric manner, implying an observation of the daily patterns of persons or groups (Craddock et al., 2016). In the absence of any formal consensus, Craddock et al. (2016) have proposed a generic set of concepts for pattern of life analysis. These concepts expand beyond the human centric context, in that they can be applied to all entities with certain behaviors. More specifically, these entities have to convey two properties, in that they must have a dynamic behavior and must be observable over time. Examples of such entities would be actions, roles, applications, processes and data points. A clear distinction is made between entities and behaviors: While entities exhibit certain patterns over time, behaviors are human understandable deductions, inferred from the behavior of entities. By this, a certain purpose can be established and linked to other observable patterns.

In recent years, the concept of pattern of life analysis has been incorporated within the field of digital forensics. Distinct from the human centric focus within intelligence communities, its primary focus is on behaviors inferred from digital activities. Such activities could be generated through user interactions and notifications, but also passively collected sensor data.

Research on pattern of life analysis gained traction in 2018 with the introduction of time-trackers in iOS and Android. Time-trackers gave rise to a new source of data, containing granular information related to device usage. With this new source of data, investigators were now able to synthesize users habits and behaviors, and to reconstruct scenarios a particular points of time. Not only this, it could be used to demonstrate premeditated actions, establish ownership of devices, and be used to support other overarching evidence (Meconi and Henseler, 2022).

In 2019, the diversity of pattern of life analysis was addressed by digital forensics researcher Sarah Edwards through a series of presentations. The presentations centered around her newly released tool APOLLO, and the tool's applicability to a wide range of scenarios. Among other things, the scenarios included analysis of health data, application usage, CarPlay activities, network usage, location data, and device lock statuses. As APOLLO combined data from a series of SQLite-databases, it managed to produce a timeline of activities with an unparalleled granularity. Edwards' presentations at various digital forensics conferences helped to further incorporate the concept of pattern of life analysis within the field of digital forensics (Edwards, 2019, 2020a,b,c).

2.5 Data at the activity data

Although no formal definition could be found in the context of digital forensics, Dr. Hans Henseler refers to data at the activity level as traces that can be used to infer or reconstruct events, thereby providing insight beyond physical evidence itself (Henseler, 2020). This aligns with a systematic review by Smit et al. (2018), whereby activity level is used to describe hypotheses about evidence that address real world actions such as violence and abuse. The term activity level is therefore closely aligned with the definition given of pattern of life analysis in the previous section. However, the term better suits the description of singular traces or events, rather than the more encompassing concept of pattern of life analysis.

Research within the digital forensic community is predominantly shared through informal platforms such as conferences, courses, blogs, forums, and other communications channels. As a consequence, there is a limited corpus of academic papers dedicated to artifacts at activity level. However, one area that has received some coverage is the reliability of artifacts within the Apple health application and traces generated by Apple CarPlay.

J. P. v. Zandwijk and Boztas (2019) performed a series of experiments to measure the accuracy of data recorded through Apple's health application. The experiments did not only focus on the accuracy of recorded steps, but included measurements of distance traveled and climbed floors. In one of the experiments, five users were equipped with iOS devices, as their walking distance, speed and carrying location were documented. The results showcased that Apple's health application recorded steps in small increments with an average error rate of about `2-3%`. The walking distance on the other side, had a deviation that reached as high as 30-40 `%` dependent on the users speed, stride length and carrying location. Despite some major deviations in recorded distance, the study concluded that the number of recorded steps can be used as s reliable source of evidence.

In another study, J. P. v. Zandwijk and Boztas (2021) illustrated that even third party communication applications such as WhatsApp can reveal user activities in the real world. In this particular instance, information about the users movements could be inferred from log files within the WhatsApp application folder. In addition, the study showcased that information about velocity and movement type could be inferred through a series of probability statements in the database `cache_encryptedC.db`, an additional source of information from Apple's health application.

Apple CarPlay was introduced in 2014, and has since become one a commonplace vehicle assistant application. As of may 2021, 80% new cars solid in the United States had support for CarPlay. Apple Carplay integrates iOS to the cars infotainment system, allowing users to place phone calls, perform messaging, play music etc. In an experiment by You et al. (2022), a series of user actions were performed on a vehicle with Apple CarPlay while driving to a specific destination. Among other things, the actions included phone calls, messaging, and the usage of Apple's voice assistant Siri. By combining information from the APOLLO tool and information from various information property lists, the researchers were able to obtain information about the vehicle name and model, Bluetooth information, application history, USB connections, details location data, text messages and more. The sum of the available information allowed for an accurate reconstruction of events taken place during the experiment.

2.6 Pattern of life analysis in criminal investigations

Despite a scarcity of academic research targeting activity related artifacts, there was no shortage of examples demonstrating its real world applicability.

At the The Digital Forensic Research Workshop in 2020, Dr Hans Henseler presented a real-world investigation, referred to as ”Murder on the Bûterwei”, to demonstrate the potential of digital traces at activity level. He emphasized how digital forensic investigators have shifted their focus from merely determining the source of a trace, to question why and how digital traces relate to a crime. He argues that this shift in focus better equip investigators to construct accurate narratives that can be evaluated against other evidential items. By utilizing the 7 w-questions, ”What happened?”, ”Where did it happen?”, and so on, investigators can create cohesive narratives and identify gaps and inconsistencies to identify areas that require further investigation (Henseler, 2020).

Examples of pattern of life analysis were found in three Norwegian court decisions. Common to all cases was its use to support overarching cohesive narratives. As the decisions revealed, the usage cases varies from supporting other circumstantial evidence to being a main evidential theme due to lack of other direct evidence.

In 2021, a woman was sentenced to one year and eight months in prison for three cases of attacks against democracy, three instances of threats and two instances of arson. One of the case's core questions was whether the defendant herself had orchestrated a series of actions, including vandalism, arson, threatening letters, to place fear in particular individuals and to claim victimhood herself. The case was highly dependent on circumstantial evidence gathered from digital devices, including her mobile phone, iPad, computer, alarm system and video surveillance system. As the prosecution lacked direct evidence, a comprehensive timeline were presented and used to scrutinize the defendant's claims and to align her movements with the events in question. A central element of the time line, was the activity data from her iPhone, which included health data, charging statuses, backlight status, application usage, and web history (TOSLO-2020-20518-2).

In 2022, a man was sentenced to 45 day in prison for neglectful driving: The defendant turned left in a intersection on a red light, hit a a scooter, resulting in the victims instantaneous death. In court, the defendant explained that while approaching the intersection, he observed a red light but believed it to have change to green. The prosecution presented a detailed timeline of events on the defendant's phone, scrutinize his actions leading up to the accident. Among other thing, the timeline illustrated that he had placing a phone call that lasted for 22 seconds, and that he had opened an email applications twice, the latter time, only 13 seconds before the incident (TOSL-2022-14854).

In another case from 2022, a man was sentenced to two years and four months of prison for aggravated robbery, unlawful detention and theft. Several men entered the victim's apartment, held a gun to her head, and locked her in a bathroom while restrained with zip ties. Among the central evidence, was a WhatsApp conversation between the victim and one of the robbers, planning to meet at the victims apartment. Information from cell phone providers tied the WhatsApp user to an iPhone 8, that was later seized from the defendant. Cell phone data also correlated iPhone 8 to the defendant's primary device, an iPhone 11, as the same sim card had been used interchangeably between both devices. Cell phone tower data places both devices in the vicinity of the crime scene shorty before the robbery. In addition, both devices were registered with detach messages only seconds apart. On the seized devices, digital forensics investigators found traces of the WhatsApp conversation, suspicious web searches included the victims address, and a significant period of inactivity was demonstrated on both devices during the time of the robbery. Location data from the iPhone 8 was further correlated with toll road records for the defendant's car, illustrating movement towards the victims residence. Investigators also demonstrated that the iPhone 8 had been connected to the defendant iPhone 11 through its wireless hot-spot functionality the same day as the robbery (LB-2022-17642).

The use of pattern of life analysis also gained mainstream recognition in the United States, when South Carolina attorney Alex Murdaugh received a life sentence for the murder of his wife and son. As the court case was broadcast to viewers around the world, the audience was given an insight into the potential use cases of traces at activity level. In a witness statement, SLED Special Agent Peter Rudofski presented a condensed timeline that included activity data from the cell phones of Alex Murdaugh, his wife Maggie and sons Paul and Buster (Rudofski, 2023, WLTX, 2023). Key aspects of the timeline included the fact that both Maggie and Paul's cell phones were locked simultaneously, only moments before they were killed. The continuation of recorded steps on Maggie's phone indicates that someone had taken her device following the death, and device orientation changes indicate that someone attempted to dispose of the device. Contrary to Alex's statements, cell tower records and a Snapchat video placed Alex in the vicinity of the crime scene only moments before the murders. Based on the detailed timeline, the prosecution argued that Alex had manipulated the crime scene in order to create an alibi (Ibbetson, 2023). In addition to these central events, the timeline also included device events such as location history, backlight status, battery level indicators, Wi-Fi connectivity, notification logs, application usage and device power events, all used within the greater narrative (Rudofski, 2023).

As claimed by Henseler (2020), the focus withing digital forensics has changes from only considering the source of a trace, to also include the context by which traces are created. Dr Henseler's statements align with the real-life cases presented in the above. In his own example, the case of the "murder on the Bûterwei", device activities were used to scrutinize the defendant's movements, and additional traces such as the device's battery, were used to contradict her statements during the interviews (Hensler, 2022). In a court case related to attacks against democracy, activity data from multiple sources were compiled into a timeline. The timeline was central to aligning the defendant's movements with events central for the case (TOSLO-2020-20518-2). When investigating a fatal car accident, the defendant's mobile device was used to establish a timeline of activities, shedding light on its usage only moments before the accident (TOSL-2022-14854). In a robbery, investigators examined device usage and location history, linking the defendants device to an iPhone 8 used in relation to the robbery (LB-2022-17642). Last, but not least, an analysis of events from multiple sized devices were central to build a timeline that played a central role in the sentencing of Alex Murdaugh for the murder of his wife and son (Rudofski, 2023).

As demonstrated, data acquired from mobile devices already play a crucial role in modern day investigations. A factor common to the cases mentioned above, is the use of digital evidence at activity level, such as device lock status, application usage, screen orientation, battery status, and more. These traces are not merely been presented as singular evidential items, but used to support overarching hypotheses.

2.7 Related work

The research question of the thesis revolves around the intersection between how tools present their data, and how this affect the human interpreter. In order gain some background knowledge related to the core research question, some related works are presented.

Itiel E Dror (2020) has performed relevant research regarding biases in expert decision making processes. A framework consisting of eight sources of biases placed within three categories are presented. These categories include biases related to specific cases (Category A), biases that are related to the person performing the analysis (Category B), and biases that are related to human nature (Category C). Several sources of biases could have relevance for the research questions of this thesis. Among other things, Category A includes biases that arise due to contextual information, such as superfluous information. Although the way in which tools presents their data is not explicitly mentioned in the study, biases due to descriptive information in digital forensic tools could have similar implications. While the study offers a comprehensive review of biases in forensic expert decision making, factors in the framework do not mention the nuance of the research question in this thesis.

I. E. Dror and Mnookin (2010) have performed interesting research regarding biases that could occur when using the Automated Fingerprint Identification Systems (AFIS). The study's core findings suggest that even though the system is meant to reduce biases, new biases are introduced through technology. The study concludes that human strategies have yet to adapt to the technological transformation in fingerprint analysis. Among the findings, the following sources of biases could have relevance to the research question of the thesis:

- Different data interpretation methods were found to affect outcomes
- The vast size of the AFIS database increased the probability of incidentally finding similar fingerprints
- Human biases occurred in relation to the systems ranking of potential fingerprint matches

The study found that differences in data interpretation methods amongst experts could lead to differences in outcome. If generalized, this source of bias could be relevant to the study, as it illustrates how use of the same tool can lead to different results. Also of interest was the increased probability of incorrect attribution when looking at large datasets in the AFIS system. The phenomenon could be of relevance as practitioners could make false attributions to data due to covariance of artifacts in granular datasets. Human biases due to the ordering of items in a list were uncovered. This phenomenon is described further in Itiel E. Dror et al. (2012), showcasing that latent fingerprint examiners were more likely to perform erroneous identifications in the top of the list, even when the correct match was presented further down the list. This phenomenon is directly relates to human interpretation of the tools presentation of data, explicitly focusing on the ordering of data.

Erlandsen (2019) focuses on fallacies in interpretation of digital evidence among prosecutors in the Norwegian police. Several inconsistencies in both identifying and weighting digital evidence were found. The study illustrated a high degree of trust in automated forensic tools, with 86 % of the respondents trusting the evidence without questioning its validity. Erlandsen's master's thesis implements a similar methodology to this study. The respondents were presented with fictitious events, followed by small excerpts of evidential data used for evaluation. However, rather than evaluating if the respondents misinterpreted the presentation of the data, it focuses on how they weight the evidential value and validity of the data.

3 Method

The chapter starts by detailing the research methodology and research design of the study. It then accounts for some of the study's delimitations, before outlining the sampling process and its limitations. The main part of the chapter is devoted to the technical aspects of the study, including hardware and software choices, in addition to various decisions regarding data reduction and normalization. Due to its comprehensive nature, creation of the scenario, including assessments given to the participants, is presented in its own chapter (4). The chapter continues by describing the analysis process, how the qualitative themes of the study are extracted, and how the quantitative aspects are measured. Finally, my own role and the preconceptions are accounted for, followed by an evaluation of the study's strengths and weaknesses.

3.1 Research methodology and philosophy

The research topic touches upon the interplay between technical aspects such as the tools presentation of data, coupled with investigator opinions, which lie within the realm of psychological factors such as biases. With an interplay of different factors at interest, a mixed-methods approach was chosen. This approach combines both qualitative and quantitative research methods, integrating findings from both methodologies into a whole. The benefit of this research methodology is that it enables researchers to gain a more comprehensive understanding of a subject, compared to a strictly qualitative or quantitative design. Instead of singling out one distinct method, the combination of methods works complements each other to gain a more comprehensive understanding (Leedy, 2021, pp. 291–293).

More specific, the study implements an explanatory sequential design. This methodology typically employs the collection and analysis of quantitative data, followed by a qualitative phase. An example of this would be the collection of quantitative data through an experiment, followed by interviews whereby participants are asked to elaborate on their experiences. This sequence of methodologies can help researchers to gain more insight into the collected data (Leedy, 2021, pp. 296–297).

The explanatory sequential design was implemented as follows: The quantitative aspect of the study seeks to measure variations in investigator conclusions, when interpreting the same scenario with results from two distinct forensic analysis tools. The qualitative aspect aims to explore the respondents experiences, especially focusing on factors within the presentation layers and whether such factors affected their interpretations. The qualitative dimension is explored with semi-structured interviews, subsequent to completing the participant assignments.

The philosophical standpoint behind the choice of research methodology is that of pragmatism, implying a belief that both absolute objective truths as well as subjective interpretations of such truths are legitimate research topics (Leedy, 2021, pp. 456). This standpoint acknowledges the complexity of the research problem in a real-world scenario, in that both the data presented through the presentation layers and the subjective experiences it generates play an important role for the outcome. Combining elements from both the tool and the human interpreter thus has the potential for greater insight than studying each in isolation.

3.2 Research design

The study's research design incorporates seven main stages: the creation of a close-to-realistic criminal incident that acts as the ground truth, acquisition of data from test devices, data reduction, processing acquired data with analysis tools, post-processing and normalization of the data, participant assessments and post response interviews, and finally an analysis and discussion of the collected data.

1. **Creation of a scenario:** This stage involves the creation of a close-to-realistic criminal incident, as to generate test data on an Apple iPhone paired with an Apple Watch. All

actions are documented to act as a ground truth when evaluating participant responses. As the scenario mimics a real-world incident, a list of investigative information needs are created. These information needs act as a basis for the participants assignments. Details around the creation of the scenario are elaborated in Chapter 4

2. **Acquisition of data from the test device:** As the above-mentioned scenario was created using an Apple iPhone paired with an Apple Watch, gaining access to their data was essential. The iPhone was exploited with a jailbreak, in order to acquire a full filesystem acquisition using openly available methods. A further elaboration of the extraction method is detailed in Section 3.9.

3. **Data reduction:** The majority of activity data stored on the iOS version used in the scenario resides within SQLite databases. As APOLLO only support parsing data from such databases, a script was used to extract all SQLite databases from the acquired full filesystem. This did not only ensure compatibility with both tools, but focused the data on activity related activities. Further details of the process can be seen in Section 3.10.

4. **Processing of Data:** The acquired data was processed with two different analysis tools, namely Cellebrite Physical Analyzer and the open-source tool APOLLO. The criteria for tools selection and the processing of data are detailed in Section 3.5.

5. **Post-processing and normalization of data:** After processing the data, some data reduction, post-processing and normalization were necessary. The reasoning behind this, was to ensure a coherent format between participant groups. Details of the process, as well as its justification is explained in Section 3.10 and 3.11.

6. **Collection of participant responses and post response Interviews:** This stage involves the collection of participant responses through a mix-methods approach. Firstly, the participants were divided into two groups, Cellebrite and APOLLO, and tasked to perform a series of assessments. Secondly, a post-scenario interview were conducted to collect subjective viewpoints from the respondents. Details of the assessments and interviews is detailed in Section 4.3 and 4.4.

7. **Analysis and discussion:** Participants responses were analysed in accordance with the research methodology. As the methodology utilized a mixed methods approach, the quantitative and qualitative analysis were combined to achieve a greater understanding of the research problem. Factors of the presentation layers, found to affect the participants, were synthesized and measured based on their occurrences. The analysis and discussion is presented in Chapter 5.

3.3 Delimitations

A distinction has been made between content data and metadata. Content data refers to elements such as files, notes, and messages, while metadata refers to descriptive information about such elements, in addition to device usage patterns, health data, location history and do on. While the assessment requires the respondents to identify some content data in form of communications (4.3), the focus is on how forensic tools present the latter category of data. As such, the respondents were not given access to the raw data source, but were confined to Excel spreadsheets with results exported from each tool.

Through the normalization process (3.11), the visual and structural effects of the tools' presentation layers was removed. The main purpose of this was harmonize the formatting of results between Cellebrite and APOLLO, in a format commonly used in forensic reports. Despite the removal of these features, they are not part of core research problem, which focuses on raw elements such as details and naming conventions within the tools presentation layer.

3.4 Sampling

Respondents for the study were selected using a purposive sampling method. This methodology entails a selection of individuals who already possess an in-depth knowledge of a certain topic. By selecting participants with a certain expertise, one can gain a more meaningful insight into a research topic (Leedy, 2021, pp. 272–273).

In the case of this study, there was a need for participants with specific knowledge within digital forensics. By this, the participants needed prior knowledge of mobile phone analysis, especially in relation to evidence at activity level. As the research topic explores a narrow field within digital forensics, recruitment of such individuals could be challenging. Therefore, the methodology of snowball sampling was used in conjunction with the purposive sampling. The snowball method implies that existing qualified participants are asked to refer to other individuals that also possess competency about a specific subject (Leedy, 2021, pp. 272–273).

By utilizing both purposive and snowball sampling in conjunction, six participants were recruited from the same police district. Despite that the participants had varying degrees of experience, all had prior knowledge of mobile phone analysis, including Cellebrite Certified Physical Analyst (CCPA) certifications.

Although a larger sample group would have provided a more extensive coverage, the number of participants was deemed satisfactory as a first outlook into the problem statement. Additional considerations were also the comprehensive nature of the data collection stage, including both quantitative data through assessments, and qualitative aspects through post-scenario interviews.

3.5 Tool selection

As described in Section 1.1, one of the driving motivations behind the thesis rests on an experience based concern regarding how digital forensic tools present their data. More specifically, if factors within the presentation layers of forensic tools, such as data normalization, affect investigators' ability to perform sound pattern of life analysis.

The de facto mobile analysis tool within Norwegian law enforcement, Cellebrite Physical Analyzer, was raised as a concern as it implements strict data normalization. The tool commonly presents its results in a fixed set of categories, within a fixed set of columns, often merging data from different sources together. Although other proprietary forensics tools share similar characteristics, the tool is of special interest due to its widespread usage.

In order to carry out the comparative analysis, there was a need to find a tool with similar capabilities of interpreting data at the activity level, but with different characteristics in its presentation layer. Two open-source forensics tools were evaluated: iLEAPP, developed by Alexis Brignoni (Brignoni, 2023), and APOLLO (Apple Pattern of Life Lazy Output'er), developed by Sarah Edwards (Edwards, 2020a).

Both tools, iLEAPP and APOLLO, were found to have advantages and disadvantages. On one hand, iLEAPP implements a modular framework that allows for custom Python modules to parse artifacts in a variety of formats. APOLLO on the other hand, implements a modular framework, but has limitations as it only parses data from SQLite databases. While iLEAPP is actively maintained, APOLLO has not received any new updates since the release of version 1.4 in December 2020. Despite APOLLO's limitations and lack of maintenance, the project was chosen due to its ability to present granular data, with no inherent limitations of categories or attributes, a radically different approach compared to Cellebrite.

The tool selection process resulted in two digital forensics tools, both parsing activity related artifacts, but with different approaches to presenting this data. Cellebrite, on one hand, enforces strict data normalization, while APOLLO provides a much more raw view of its data. These different approaches are believed to provide the opportunity to differentiate participant evaluations based on their different approaches for presentation of data.

3.6 Hardware

The scenario was dependent on an Apple iPhone paired with an Apple Watch for data generation. Due to security measures implemented in the devices, the user is by default denied access to most databases containing activity data. To circumvent these security measures, an Apple iPhone capable of being jailbroken was necessary. As Apple Watches are automatically synchronised with its parent device, no separate data extraction is necessary.

Apple iPhone SE (2. gen) with iOS 14.8:
The amount of data that can be parsed by digital forensic tools is entirely dependent on the data acquired from a device: while a standard iTunes backup gives access to content like communications and media files, a full filesystem extraction includes most files, including essential system database. However, to gain access to the required files with publicly available methods, it was necessary to perform a jailbreak on the test device.

The process of jailbreaking an iOS device entails the removal of several security features enabled by Apple. Among other things, it allows a user to install and run unsigned applications and to gain root access to a device.[1] However, the availability of jailbreaking methods is highly dependent on the iPhone model and installed version of iOS.

Two different jailbreak methods were evaluated, namely Palera1n[2] and Unc0ver.[3] Among these, Palera1n is the most recent, with support up to iOS 16. Despite having the most recent support, it is necessary to the device's lock code for the jailbreak to function. Unc0ver, on the other hand, supports iOS version up until 14.8, and allows the user to maintain a lock code on the device. Common for both methods, is that only older models of iPhone are supported, with iPhone SE (2. gen) among the last supported devices.

With an iPhone SE (2. gen) running iOS 14.8 readily available, the device was chosen as it met all the necessary requirements. It was suitable for jailbreak with Unc0ver, allowed the lock code to be maintained, and has an iOS version that is supported by both Cellebrite and APOLLO. Although newer versions of iOS is available, version 14.8 is still in active use. Therefore, generating data through the selected device is still of relevance.

Apple Watch SE:
As the thesis focuses on analysis of activity data, including health data, an Apple Watch SE was added to the scenario. The initial plan was to perform a file system extraction on the Apple Watch itself. However, testing revealed that all required health data were automatically synchronized with the paired iPhone, deeming a separate extraction unnecessary. The main reason for selecting the Apple Watch SE was its availability at the time of the scenario.

3.7 Pilot study

Prior to the main data collection stage, a pilot study was conducted. The purpose of the pilot study was to gain experience with the validity and efficiency of the data collection methods, and to fine-tune the information needs and post-scenario interviews. The pilot study was carried out by an individual that did not take part in the study, measuring the time used for each individual assignment. Following the pilot study, each stage of the data collection process was evaluated and adjusted accordingly.

Several challenges were identified during the initial data collection process. For practical reasons, the time needed for each participant to complete their assignments had to be limited to approximately 60 minutes. The pilot study revealed that several information needs were overly time-consuming, and that others were repetitive, involving nearly identical operations. Additionally, some tasks were reformulated, and others rearranged or merged together.

[1] https://www.mcafee.com/blogs/mobile-security/how-does-jailbreaking-or-rooting-affect-my-mobile-device-security/

[2] https://palera.in

[3] https://unc0ver.dev

A similar process was implemented for the post-scenario interview. During the practice run, it became clear that some of the questions were somewhat unclear and needed to be reformulated. Others needed to be rearranged to improve the flow of the questionnaire, starting with the more open-ended questions and gradually working towards more specifics.

The pilot study resulted in an almost 50% reduction in assessment time. Although necessary to uphold the time limitations, several important information needs were eliminated, somewhat limiting the participants ability to gain a complete overview of the scenario.

3.8 Data creation

In order to establish a ground truth for evaluation purposes, a controlled experiment was conducted. The experiment included the creation of a close-to-realistic criminal incident, to generated data for the ground truth. During the data creation, a test subject performed a series of pre-arranged actions, designed to simulate the murder of a victim. These actions included physical activities like walking, running, climbing up stairs, as well as interactions with the device, such as unlocking the phone, sending text messages and connecting the device to different peripherals. As to measure the participants assessments, each action in the experiment was documented. The resulting timeline functions as a baseline of ground truth for evaluation of participant responses. A detailed account of the data creation phase can be found in Chapter 4.

With regards to the methodology used to create the scenario, the procedures were inspired by Joshua Hickman's publicly available test images. Joshua Hickman is currently employed as a Subject Matter Expert in the Collect & Review LOB at Cellebrite, and is a leading contributor within the digital forensic community[4]. As of my knowledge his test procedures is not grounded in any specific theoretical framework but is highly regarded within the community. Detailed information about the creation of Hickmans's test images is embedded as PDF files within test images found on thebinaryhick.blog (Hickman, 2021)

3.9 Data extraction

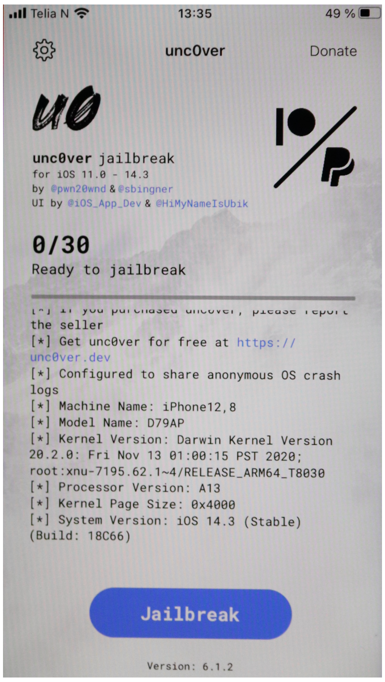
To extract data from the iPhone following the scenario, it was necessary to gain root access to the device. The section proceeds to showcase the jailbreak process, including establishing SSH root access and the usage of iOS Triage for a full filesystem extraction.

Jailbreak with Unc0ver:
As discussed in Section 3.6, an Apple iPhone SE (2. gen) with iOS 14.8 was selected, as it is supported by the Unc0ver jailbreak and has an iOS version that is parsed by both APOLLO and Cellebrite.

Two methods for deploying Unc0ver were evaluated, side loading Unc0ver with AltStore and installing the Unc0ver through 3uTools. Of the two methods, 3uTools facilitated a one-click approach and was therefore selected over AltStore.

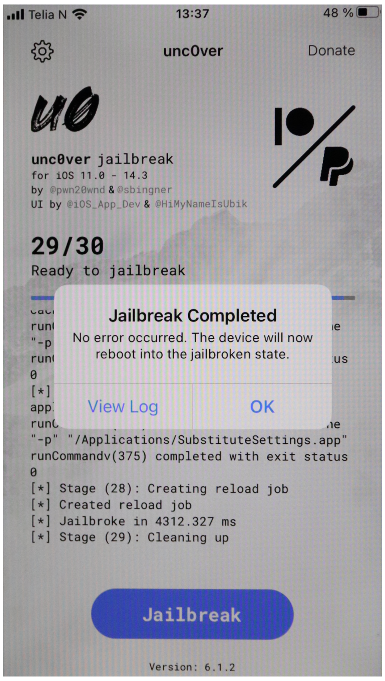
Unc0ver was installed on the device through 3uTools. After the initial reboot, Unc0ver was successfully installed. To verify that the phone was successfully jailbroken, it was connected to a MacBook Pro through a lightning cable. SSH was launched from the terminal and after accepting the phone's SSH certificate, a successful SSH connection was established. A listing of the device's file structure confirmed root access to the device.

[4] https://www.sans.org/profiles/joshua-hickman



```
[*] If you purchased unc0ver, please report
the seller
[*] Get unc0ver for free at https://
unc0ver.dev
[*] Configured to share anonymous OS crash
logs
[*] Machine Name: iPhone12,8
[*] Model Name: D79AP
[*] Kernel Version: Darwin Kernel Version
20.2.0: Fri Nov 13 01:00:15 PST 2020;
root:xnu-7195.62.1~4/RELEASE_ARM64_T8030
[*] Processor Version: A13
[*] Kernel Page Size: 0x4000
[*] System Version: iOS 14.3 (Stable)
(Build: 18C66)
```

Figure 1: Unc0ver jailbreak loaded on device

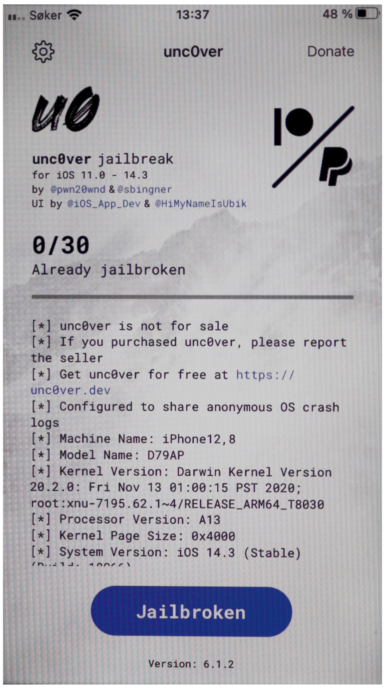


Jailbreak Completed

No error occurred. The device will now reboot into the jailbroken state.

```
runC                                        ne
"-p
runC                                       tus
0
[*]
app
runC                                        ne
"-p" "/Applications/SubstituteSettings.app"
runCommandv(375) completed with exit status
0
[*] Stage (28): Creating reload job
[*] Created reload job
[*] Jailbroke in 4312.327 ms
[*] Stage (29): Cleaning up
```

Figure 2: Unc0ver jailbreak process completed



```
unc0ver jailbreak
for iOS 11.0 - 14.3
by @pwn20wnd & @sbingner
UI by @iOS_App_Dev & @HiMyNameIsUbik

0/30
Already jailbroken

[*] unc0ver is not for sale
[*] If you purchased unc0ver, please report
the seller
[*] Get unc0ver for free at https://
unc0ver.dev
[*] Configured to share anonymous OS crash
logs
[*] Machine Name: iPhone12,8
[*] Model Name: D79AP
[*] Kernel Version: Darwin Kernel Version
20.2.0: Fri Nov 13 01:00:15 PST 2020;
root:xnu-7195.62.1~4/RELEASE_ARM64_T8030
[*] Processor Version: A13
[*] Kernel Page Size: 0x4000
[*] System Version: iOS 14.3 (Stable)
```

Figure 3: Device successfully jailbroken with Unc0ver

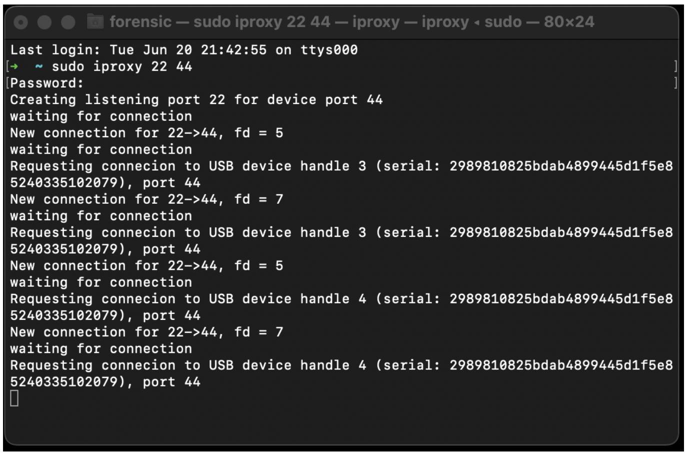
Data extraction with RealityNet iOS Triage:
With root access to the device, files could be transferred successfully with SCP. Despite that SCP allowed for transfer of single files, an alternate method of acquiring the entire file system was needed.

A search for openly available acquisition methods revealed a tutorial from the Sans institute[5] referencing the open source script iOS Triage.[6] iOS Triage is a Bash script created by Mattia Epifani that, among other things is capable of performing a full filesystem extraction from jailbroken iOS devices.

When following the tutorial given on the project's GitHub page, it became clear that a series of preparatory steps were necessary. Firstly there was a need to install dependencies and grant execution privileges to the bash script. Secondly iproxy needed to be set up to allow the script access to the device.

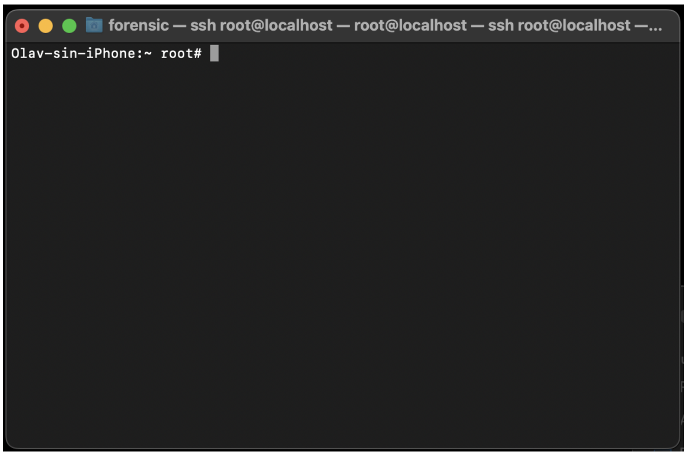
With the preparatory steps completed, the bash script was executed on the MacBook Pro. iOS Triage was presented by a easy to use menu, whereby the choice for a full filesystem extraction was included. Executing this option resulted in a complete acquisition of the device's filesystem, contained within a single tar archive.

[5] https://www.sans.org/blog/checkra1n---part-1---prep
[6] https://github.com/RealityNet/ios_triage


forensic — sudo iproxy 22 44 — iproxy — iproxy ◂ sudo — 80×24

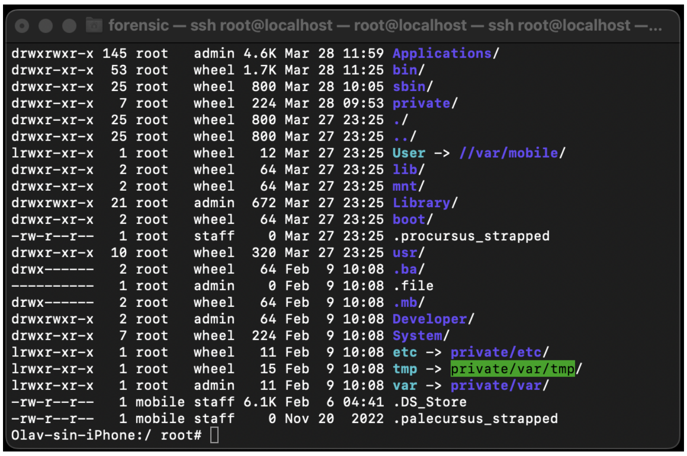
```
Last login: Tue Jun 20 21:42:55 on ttys000
→  ~ sudo iproxy 22 44
Password:
Creating listening port 22 for device port 44
waiting for connection
New connection for 22->44, fd = 5
waiting for connection
Requesting connecion to USB device handle 3 (serial: 2989810825bdab4899445d1f5e8
5240335102079), port 44
New connection for 22->44, fd = 7
waiting for connection
Requesting connecion to USB device handle 3 (serial: 2989810825bdab4899445d1f5e8
5240335102079), port 44
New connection for 22->44, fd = 5
waiting for connection
Requesting connecion to USB device handle 4 (serial: 2989810825bdab4899445d1f5e8
5240335102079), port 44
New connection for 22->44, fd = 7
waiting for connection
Requesting connecion to USB device handle 4 (serial: 2989810825bdab4899445d1f5e8
5240335102079), port 44
```

Figure 4: iProxy started on port 22 44


forensic — ssh root@localhost — root@localhost — ssh root@localhost —...

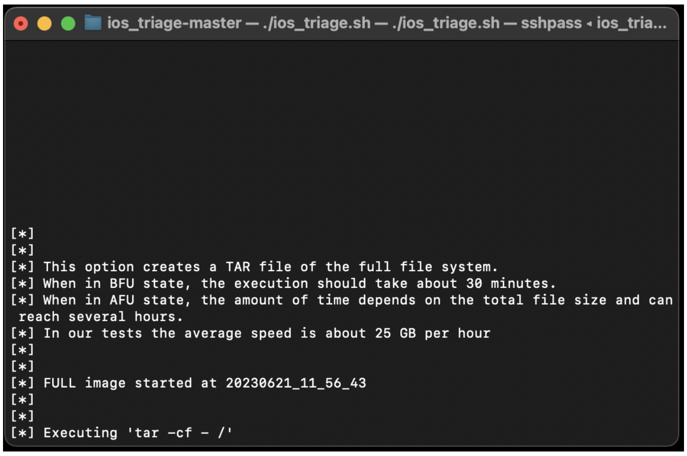
```
Olav-sin-iPhone:~ root#
```

Figure 5: SSH root access to the device


forensic — ssh root@localhost — root@localhost — ssh root@localhost —...

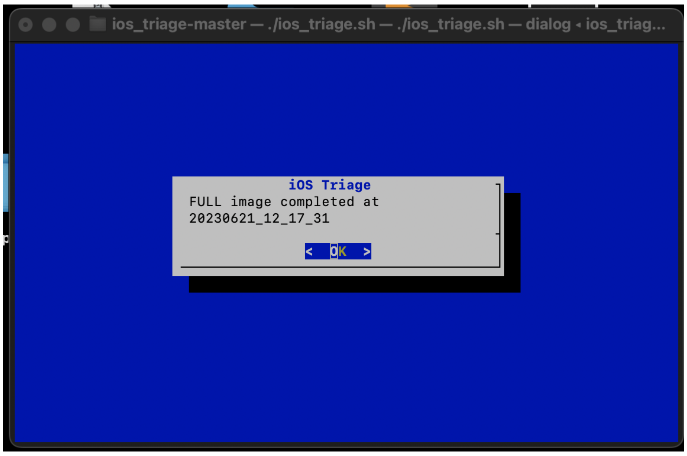
```
drwxrwxr-x 145 root   admin 4.6K Mar 28 11:59 Applications/
drwxr-xr-x  53 root   wheel 1.7K Mar 28 11:25 bin/
drwxr-xr-x  25 root   wheel  800 Mar 28 10:05 sbin/
drwxr-xr-x   7 root   wheel  224 Mar 28 09:53 private/
drwxr-xr-x  25 root   wheel  800 Mar 27 23:25 ./
drwxr-xr-x  25 root   wheel  800 Mar 27 23:25 ../
lrwxr-xr-x   1 root   wheel   12 Mar 27 23:25 User -> //var/mobile/
drwxr-xr-x   2 root   wheel   64 Mar 27 23:25 lib/
drwxr-xr-x   2 root   wheel   64 Mar 27 23:25 mnt/
drwxrwxr-x  21 root   admin  672 Mar 27 23:25 Library/
drwxr-xr-x   2 root   wheel   64 Mar 27 23:25 boot/
-rw-r--r--   1 root   staff    0 Mar 27 23:25 .procursus_strapped
drwxr-xr-x  10 root   wheel  320 Mar 27 23:25 usr/
drwx------   2 root   wheel   64 Feb  9 10:08 .ba/
----------   1 root   admin    0 Feb  9 10:08 .file
drwx------   2 root   wheel   64 Feb  9 10:08 .mb/
drwxrwxr-x   2 root   admin   64 Feb  9 10:08 Developer/
drwxr-xr-x   7 root   wheel  224 Feb  9 10:08 System/
lrwxr-xr-x   1 root   wheel   11 Feb  9 10:08 etc -> private/etc/
lrwxr-xr-x   1 root   wheel   15 Feb  9 10:08 tmp -> private/var/tmp/
lrwxr-xr-x   1 root   admin   11 Feb  9 10:08 var -> private/var/
-rw-r--r--   1 mobile staff 6.1K Feb  6 04:41 .DS_Store
-rw-r--r--   1 mobile staff    0 Nov 20  2022 .palecursus_strapped
Olav-sin-iPhone:/ root#
```

Figure 6: SSH root access to the device, list list of files


ios_triage-master — ./ios_triage.sh — ./ios_triage.sh — sshpass ◂ ios_tria...

```
[*]
[*]
[*] This option creates a TAR file of the full file system.
[*] When in BFU state, the execution should take about 30 minutes.
[*] When in AFU state, the amount of time depends on the total file size and can
 reach several hours.
[*] In our tests the average speed is about 25 GB per hour
[*]
[*]
[*] FULL image started at 20230621_11_56_43
[*]
[*]
[*] Executing 'tar -cf - /'
```

Figure 7: iOS Triage full filesystem extraction started


ios_triage-master — ./ios_triage.sh — ./ios_triage.sh — dialog ◂ ios_triag...

```
iOS Triage
FULL image completed at
20230621_12_17_31

< OK >
```

Figure 8: iOS Triage full filesystem extraction completed

3.10 Data reduction

The fictitious scenario, as described in Chapter 4, was created within a time frame of about 12 hours. Despite that the scenario included a limited set of activities, large amounts of data were generated.

The study focuses on activity data generated on iOS devices. Given that the majority of activity data on iOS 14.8 is stored within SQLite databases, a natural limitation was to only process such databases. This decision did not only ensured compatibility with both analysis tools, but concentrated the results to activity related data.

A Python script was created to recursively extract all SQLite databases from the full filesystem extraction. The script ensured that all extracted databases retained their original file structure to ensure compatibility with both tools. The reduced acquisition file was then processed with both Cellebrite Physical Analyzer and APOLLO, and result from each respective tool was exported.

The resulting timelines included activity data that spanned over several hundred thousand rows. To present a more manageable dataset, it was essential to filter out data that fell outside the time span of the scenario. The remaining data was chosen based on the natural start and end time of the scenario, which included about 12 hours of activity data.

Databases and rows for each tool:
Despite that both tools were given access to the same source material, there was a major discrepancy between the results. This discrepancy presented itself, both in the number of databases processed and the number of rows produced by each tool, factors that is dependent on the capabilities of the parsers included in each tool. After filtering out data that lie outside the span of the scenario, the tools were left with the following results: Cellebrite processed a total of 12 databases, resulting in 13,979 rows, while APOLLO processed a total of 16 databases, resulting in 55,225 rows. Of 18 processed databases in total, 10 were processed by both tools. A detailed count of rows extracted for each database is provided in Table 1 below.

Cellebrite	**Rows**	**APOLLO**	**Rows**
BrowserState.db	1	-	-
Cache.sqlite	8133	Cache.sqlite	8222
CurrentPowerlog.PLSQL	58	CurrentPowerlog.PLSQL	38841
DataUsage.sqlite	65	DataUsage.sqlite	87
cache_encryptedB.db	4840	cache_encryptedB.db	4802
consolidated.db	22	-	-
healthdb_secure.sqlite	28	healthdb_secure.sqlite	408
History.db	26	History.db	13
interactionC.db	9	interactionC.db	5
knowledgeC.db	770	knowledgeC.db	1228
Local.sqlite	22	Local.sqlite	22
sms.db	5	sms.db	8
-	-	Cloud-V2.sqlite	22
-	-	Photos.sqlite	2
-	-	RMAdminStore-Local.sqlite	78
-	-	TCC.db	1
-	-	cache_encryptedC.db	1343
-	-	netusage.sqlite	140
Total Rows	**13,979**	**Total Rows**	**55,225**

Table 1: Databases and number of rows processed by Cellebrite and APOLLO

Activities and rows for each tool:
APOLLO segmented its results within 101 unique categories. Table 2 below contains an excerpt of the top 25 activity types in APOLLO. The entirety of the table can be seen in Section 5.

Activity	Count
Device/App Assertions	29078
Routined Location	8200
WiFi Location	4602
App Usage Video	1783
Kernel Task Monitor	885
Coalition Interval	864
Battery Level	853
Health - Step Count	739
Motion State History	604
Process Data Usage	582
Screen Brightness	570
Video CM File	555
App Location Usage	544
Video VT Session	528
App Usage	396
Springboard Screen State	308
Network Usage	296
WiFi Connection	289
Health Heart Rate	268
Location Technology	238
Activity States	213
Cellular Location	200
Now Playing	169
Process ID	169
Device Volume	148

Table 2: Excerpt of top 25 activity types in APOLLO

As seen by Table 3, Cellebrite on the other hand placed its results in 14 different categories.

Activity	Count
Locations	8215
Wireless Network: Location	4573
Device Events	471
Applications Usage Log	313
Cell Tower: Location	200
Log Entries	70
Wireless Networks	29
Activity Sensor Data	28
Device Connectivity	24
Web History	21
Searched Items	20
Instant Messages	10
Contacts	4
Installed Applications	1

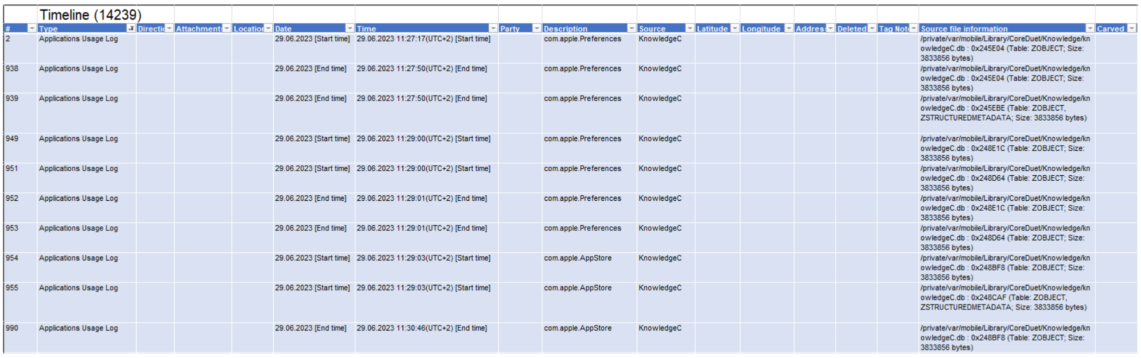
Table 3: Activity types in Cellebrite

3.11 Data processing and normalization

Following the data extraction, reduction and processing phase, it was necessary to standardize the exported results to a format consistent between both tools.

Data normalization in Cellebrite:
After processing the reduced acquisition with Cellebrite Physical Analyzer, results from the tool's

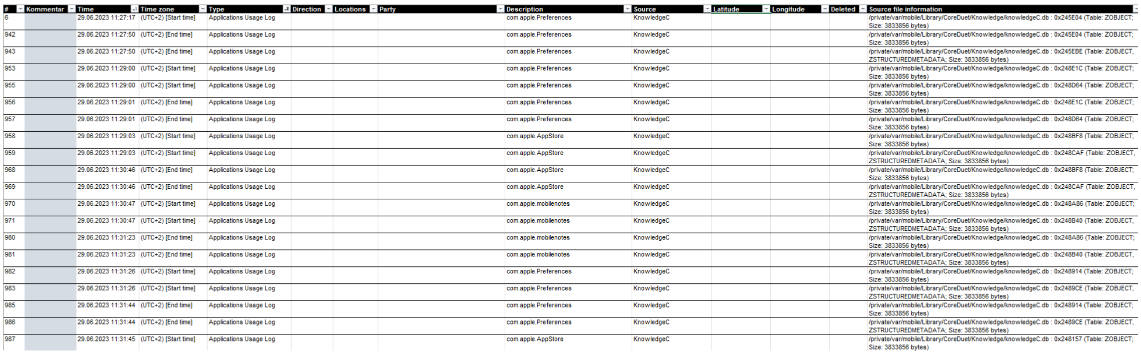
timeline were exported to an Excel spreadsheet. With inherent support for export, only minor adjustments were needed for the final result. These adjustments included an additional column for participant comments, removal of duplicate timestamp columns, rearrangement of columns, and implementing a standardized style. Table 4 and 5, illustrate the results before and after the changes.



Timeline (14239)

#	Type	Directi	Attachment	Locatio	Date	Time	Party	Description	Source	Latitude	Longitude	Addres	Deleted	Tag Not	Source file information	Carved
2	Applications Usage Log				29.06.2023 [Start time]	29.06.2023 11:27:17(UTC+2) [Start time]		com.apple.Preferences	KnowledgeC						/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x245E04 (Table: ZOBJECT; Size: 3833856 bytes)	
938	Applications Usage Log				29.06.2023 [End time]	29.06.2023 11:27:50(UTC+2) [End time]		com.apple.Preferences	KnowledgeC						/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x245E04 (Table: ZOBJECT; Size: 3833856 bytes)	
939	Applications Usage Log				29.06.2023 [End time]	29.06.2023 11:27:50(UTC+2) [End time]		com.apple.Preferences	KnowledgeC						/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x245EBE (Table: ZOBJECT, ZSTRUCTUREDMETADATA; Size: 3833856 bytes)	
949	Applications Usage Log				29.06.2023 [Start time]	29.06.2023 11:29:00(UTC+2) [Start time]		com.apple.Preferences	KnowledgeC						/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248E1C (Table: ZOBJECT; Size: 3833856 bytes)	
951	Applications Usage Log				29.06.2023 [Start time]	29.06.2023 11:29:00(UTC+2) [Start time]		com.apple.Preferences	KnowledgeC						/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248D64 (Table: ZOBJECT; Size: 3833856 bytes)	
952	Applications Usage Log				29.06.2023 [End time]	29.06.2023 11:29:01(UTC+2) [End time]		com.apple.Preferences	KnowledgeC						/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248E1C (Table: ZOBJECT; Size: 3833856 bytes)	
953	Applications Usage Log				29.06.2023 [End time]	29.06.2023 11:29:01(UTC+2) [End time]		com.apple.Preferences	KnowledgeC						/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248D64 (Table: ZOBJECT; Size: 3833856 bytes)	
954	Applications Usage Log				29.06.2023 [Start time]	29.06.2023 11:29:03(UTC+2) [Start time]		com.apple.AppStore	KnowledgeC						/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248BF8 (Table: ZOBJECT; Size: 3833856 bytes)	
955	Applications Usage Log				29.06.2023 [Start time]	29.06.2023 11:29:03(UTC+2) [Start time]		com.apple.AppStore	KnowledgeC						/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248CAF (Table: ZOBJECT, ZSTRUCTUREDMETADATA; Size: 3833856 bytes)	
990	Applications Usage Log				29.06.2023 [End time]	29.06.2023 11:30:46(UTC+2) [End time]		com.apple.AppStore	KnowledgeC						/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248BF8 (Table: ZOBJECT; Size: 3833856 bytes)	

Table 4: Excel spreadsheet exported from Cellebrite's timeline view



#	Kommentar	Time	Time zone	Type	Direction	Locations	Party	Description	Source	Latitude	Longitude	Deleted	Source file information
6		29.06.2023 11:27:17	(UTC+2) [Start time]	Applications Usage Log				com.apple.Preferences	KnowledgeC				/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x245E04 (Table: ZOBJECT; Size: 3833856 bytes)
942		29.06.2023 11:27:50	(UTC+2) [End time]	Applications Usage Log				com.apple.Preferences	KnowledgeC				/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x245E04 (Table: ZOBJECT; Size: 3833856 bytes)
943		29.06.2023 11:27:50	(UTC+2) [End time]	Applications Usage Log				com.apple.Preferences	KnowledgeC				/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x245EBE (Table: ZOBJECT, ZSTRUCTUREDMETADATA; Size: 3833856 bytes)
953		29.06.2023 11:29:00	(UTC+2) [Start time]	Applications Usage Log				com.apple.Preferences	KnowledgeC				/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248E1C (Table: ZOBJECT; Size: 3833856 bytes)
955		29.06.2023 11:29:00	(UTC+2) [Start time]	Applications Usage Log				com.apple.Preferences	KnowledgeC				/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248D64 (Table: ZOBJECT; Size: 3833856 bytes)
956		29.06.2023 11:29:01	(UTC+2) [End time]	Applications Usage Log				com.apple.Preferences	KnowledgeC				/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248E1C (Table: ZOBJECT; Size: 3833856 bytes)
957		29.06.2023 11:29:01	(UTC+2) [End time]	Applications Usage Log				com.apple.Preferences	KnowledgeC				/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248D64 (Table: ZOBJECT; Size: 3833856 bytes)
958		29.06.2023 11:29:03	(UTC+2) [Start time]	Applications Usage Log				com.apple.AppStore	KnowledgeC				/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248BF8 (Table: ZOBJECT; Size: 3833856 bytes)
959		29.06.2023 11:29:03	(UTC+2) [Start time]	Applications Usage Log				com.apple.AppStore	KnowledgeC				/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248CAF (Table: ZOBJECT, ZSTRUCTUREDMETADATA; Size: 3833856 bytes)
968		29.06.2023 11:30:46	(UTC+2) [End time]	Applications Usage Log				com.apple.AppStore	KnowledgeC				/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248BF8 (Table: ZOBJECT; Size: 3833856 bytes)
969		29.06.2023 11:30:46	(UTC+2) [End time]	Applications Usage Log				com.apple.AppStore	KnowledgeC				/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248CAF (Table: ZOBJECT, ZSTRUCTUREDMETADATA; Size: 3833856 bytes)
970		29.06.2023 11:30:47	(UTC+2) [Start time]	Applications Usage Log				com.apple.mobilenotes	KnowledgeC				/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248A86 (Table: ZOBJECT; Size: 3833856 bytes)
971		29.06.2023 11:30:47	(UTC+2) [Start time]	Applications Usage Log				com.apple.mobilenotes	KnowledgeC				/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248B40 (Table: ZOBJECT, ZSTRUCTUREDMETADATA; Size: 3833856 bytes)
980		29.06.2023 11:31:23	(UTC+2) [End time]	Applications Usage Log				com.apple.mobilenotes	KnowledgeC				/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248A86 (Table: ZOBJECT; Size: 3833856 bytes)
981		29.06.2023 11:31:23	(UTC+2) [End time]	Applications Usage Log				com.apple.mobilenotes	KnowledgeC				/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248B40 (Table: ZOBJECT, ZSTRUCTUREDMETADATA; Size: 3833856 bytes)
982		29.06.2023 11:31:26	(UTC+2) [Start time]	Applications Usage Log				com.apple.Preferences	KnowledgeC				/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248914 (Table: ZOBJECT; Size: 3833856 bytes)
983		29.06.2023 11:31:26	(UTC+2) [Start time]	Applications Usage Log				com.apple.Preferences	KnowledgeC				/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x2489CE (Table: ZOBJECT, ZSTRUCTUREDMETADATA; Size: 3833856 bytes)
985		29.06.2023 11:31:44	(UTC+2) [End time]	Applications Usage Log				com.apple.Preferences	KnowledgeC				/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248914 (Table: ZOBJECT; Size: 3833856 bytes)
986		29.06.2023 11:31:44	(UTC+2) [End time]	Applications Usage Log				com.apple.Preferences	KnowledgeC				/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x2489CE (Table: ZOBJECT, ZSTRUCTUREDMETADATA; Size: 3833856 bytes)
987		29.06.2023 11:31:45	(UTC+2) [Start time]	Applications Usage Log				com.apple.AppStore	KnowledgeC				/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x248157 (Table: ZOBJECT; Size: 3833856 bytes)

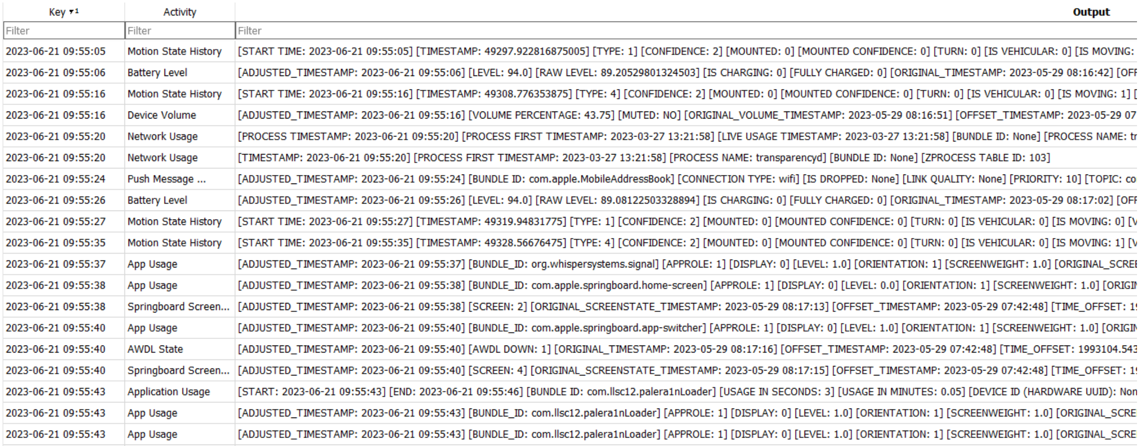
Table 5: Excel spreadsheet from Cellebrite timeline, normalized to its final form

Data normalization in APOLLO:
In contrast to Cellebrite's export, which only required minor adjustments, the results from APOLLO needed extensive post-processing to reach a comparable format.

APOLLO is a command-line tool written in Python. The tool traverses the entirety of the acquired folder structure, and extracts data from SQLite databases defined within its modules. The extracted data is compiled into a single timeline, that can be stored as a CSV file or an SQLite database. Regarding APOLLO's attributes, the data cannot be considered normalized, as the number of properties varies across different activities. The tool's solution to this is to place all available data within a single column, where each distinct property is separated by square brackets

Another issue that required attention was that all timestamps are stored in Coordinated Universal Time (UTC). To produce a format that was coherent with the results from Cellebrite, all timestamps needed conversion to Norwegian local time UTC(+2).

The full filesystem acquisition was processed with APOLLO, and its results were written to a SQLite database. The resulting SQLite database stored values within three columns: "timestamp", which displays the timestamp of the activity; "activity", which provides the name of the activity; and "output", containing key-value pairs for all of the parsed information. Table 6 illustrates how the parsed data is enclosed in square brackets within the "output" column.

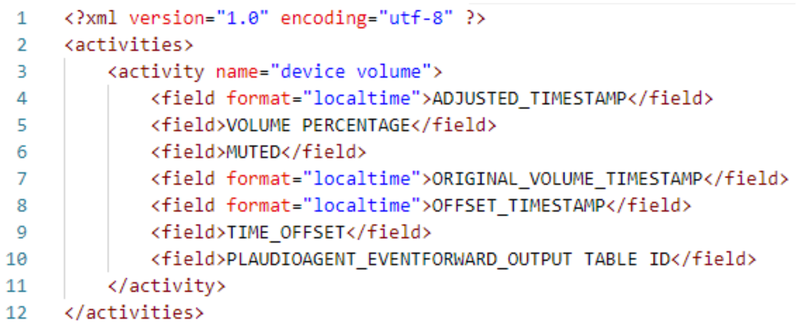


Key ▾1	Activity	Output
Filter	Filter	Filter
2023-06-21 09:55:05	Motion State History	[START TIME: 2023-06-21 09:55:05] [TIMESTAMP: 49297.922816875005] [TYPE: 1] [CONFIDENCE: 2] [MOUNTED: 0] [MOUNTED CONFIDENCE: 0] [TURN: 0] [IS VEHICULAR: 0] [IS MOVING:
2023-06-21 09:55:06	Battery Level	[ADJUSTED_TIMESTAMP: 2023-06-21 09:55:06] [LEVEL: 94.0] [RAW LEVEL: 89.20529801324503] [IS CHARGING: 0] [FULLY CHARGED: 0] [ORIGINAL_TIMESTAMP: 2023-05-29 08:16:42] [OFF
2023-06-21 09:55:16	Motion State History	[START TIME: 2023-06-21 09:55:16] [TIMESTAMP: 49308.776353875] [TYPE: 4] [CONFIDENCE: 2] [MOUNTED: 0] [MOUNTED CONFIDENCE: 0] [TURN: 0] [IS VEHICULAR: 0] [IS MOVING: 1] [
2023-06-21 09:55:16	Device Volume	[ADJUSTED_TIMESTAMP: 2023-06-21 09:55:16] [VOLUME PERCENTAGE: 43.75] [MUTED: NO] [ORIGINAL_VOLUME_TIMESTAMP: 2023-05-29 08:16:51] [OFFSET_TIMESTAMP: 2023-05-29 07
2023-06-21 09:55:20	Network Usage	[PROCESS TIMESTAMP: 2023-06-21 09:55:20] [PROCESS FIRST TIMESTAMP: 2023-03-27 13:21:58] [LIVE USAGE TIMESTAMP: 2023-03-27 13:21:58] [BUNDLE ID: None] [PROCESS NAME: tr
2023-06-21 09:55:20	Network Usage	[TIMESTAMP: 2023-06-21 09:55:20] [PROCESS FIRST TIMESTAMP: 2023-03-27 13:21:58] [PROCESS NAME: transparencyd] [BUNDLE ID: None] [ZPROCESS TABLE ID: 103]
2023-06-21 09:55:24	Push Message ...	[ADJUSTED_TIMESTAMP: 2023-06-21 09:55:24] [BUNDLE ID: com.apple.MobileAddressBook] [CONNECTION TYPE: wifi] [IS DROPPED: None] [LINK QUALITY: None] [PRIORITY: 10] [TOPIC: co
2023-06-21 09:55:26	Battery Level	[ADJUSTED_TIMESTAMP: 2023-06-21 09:55:26] [LEVEL: 94.0] [RAW LEVEL: 89.08122503328894] [IS CHARGING: 0] [FULLY CHARGED: 0] [ORIGINAL_TIMESTAMP: 2023-05-29 08:17:02] [OFF
2023-06-21 09:55:27	Motion State History	[START TIME: 2023-06-21 09:55:27] [TIMESTAMP: 49319.94831775] [TYPE: 1] [CONFIDENCE: 2] [MOUNTED: 0] [MOUNTED CONFIDENCE: 0] [TURN: 0] [IS VEHICULAR: 0] [IS MOVING: 0] [V
2023-06-21 09:55:35	Motion State History	[START TIME: 2023-06-21 09:55:35] [TIMESTAMP: 49328.56676475] [TYPE: 4] [CONFIDENCE: 2] [MOUNTED: 0] [MOUNTED CONFIDENCE: 0] [TURN: 0] [IS VEHICULAR: 0] [IS MOVING: 1] [V
2023-06-21 09:55:37	App Usage	[ADJUSTED_TIMESTAMP: 2023-06-21 09:55:37] [BUNDLE_ID: org.whispersystems.signal] [APPROLE: 1] [DISPLAY: 0] [LEVEL: 1.0] [ORIENTATION: 1] [SCREENWEIGHT: 1.0] [ORIGINAL_SCREE
2023-06-21 09:55:38	App Usage	[ADJUSTED_TIMESTAMP: 2023-06-21 09:55:38] [BUNDLE_ID: com.apple.springboard.home-screen] [APPROLE: 1] [DISPLAY: 0] [LEVEL: 0.0] [ORIENTATION: 1] [SCREENWEIGHT: 1.0] [ORIGIN
2023-06-21 09:55:38	Springboard Screen...	[ADJUSTED_TIMESTAMP: 2023-06-21 09:55:38] [SCREEN: 2] [ORIGINAL_SCREENSTATE_TIMESTAMP: 2023-05-29 08:17:13] [OFFSET_TIMESTAMP: 2023-05-29 07:42:48] [TIME_OFFSET: 1
2023-06-21 09:55:40	App Usage	[ADJUSTED_TIMESTAMP: 2023-06-21 09:55:40] [BUNDLE_ID: com.apple.springboard.app-switcher] [APPROLE: 1] [DISPLAY: 0] [LEVEL: 1.0] [ORIENTATION: 1] [SCREENWEIGHT: 1.0] [ORIGIN
2023-06-21 09:55:40	AWDL State	[ADJUSTED_TIMESTAMP: 2023-06-21 09:55:40] [AWDL DOWN: 1] [ORIGINAL_TIMESTAMP: 2023-05-29 08:17:16] [OFFSET_TIMESTAMP: 2023-05-29 07:42:48] [TIME_OFFSET: 1993104.543
2023-06-21 09:55:40	Springboard Screen...	[ADJUSTED_TIMESTAMP: 2023-06-21 09:55:40] [SCREEN: 4] [ORIGINAL_SCREENSTATE_TIMESTAMP: 2023-05-29 08:17:15] [OFFSET_TIMESTAMP: 2023-05-29 07:42:48] [TIME_OFFSET: 1
2023-06-21 09:55:43	Application Usage	[START: 2023-06-21 09:55:43] [END: 2023-06-21 09:55:46] [BUNDLE ID: com.llsc12.palera1nLoader] [USAGE IN SECONDS: 3] [USAGE IN MINUTES: 0.05] [DEVICE ID (HARDWARE UUID): Non
2023-06-21 09:55:43	App Usage	[ADJUSTED_TIMESTAMP: 2023-06-21 09:55:43] [BUNDLE_ID: com.llsc12.palera1nLoader] [APPROLE: 1] [DISPLAY: 0] [LEVEL: 1.0] [ORIENTATION: 1] [SCREENWEIGHT: 1.0] [ORIGINAL_SCRE
2023-06-21 09:55:43	App Usage	[ADJUSTED_TIMESTAMP: 2023-06-21 09:55:43] [BUNDLE_ID: com.llsc12.palera1nLoader] [APPROLE: 1] [DISPLAY: 0] [LEVEL: 1.0] [ORIENTATION: 1] [SCREENWEIGHT: 1.0] [ORIGINAL_SCRE

Table 6: APOLLO encloses attributes within square brackets

A Python script was developed to address the challenges related to attributes enclosed within square brackets. Content within the ”output” column was split into a list of distinct key-value pairs based on the separator. The key-value pairs were then transposed, such that each key-value pair was shown as a separate lines within the output column. The Python script also converted all timestamps from UTC to Norwegian local time, regardless if they were contained in the ”key” column, or embedded within key-value pairs in the ”output” column.

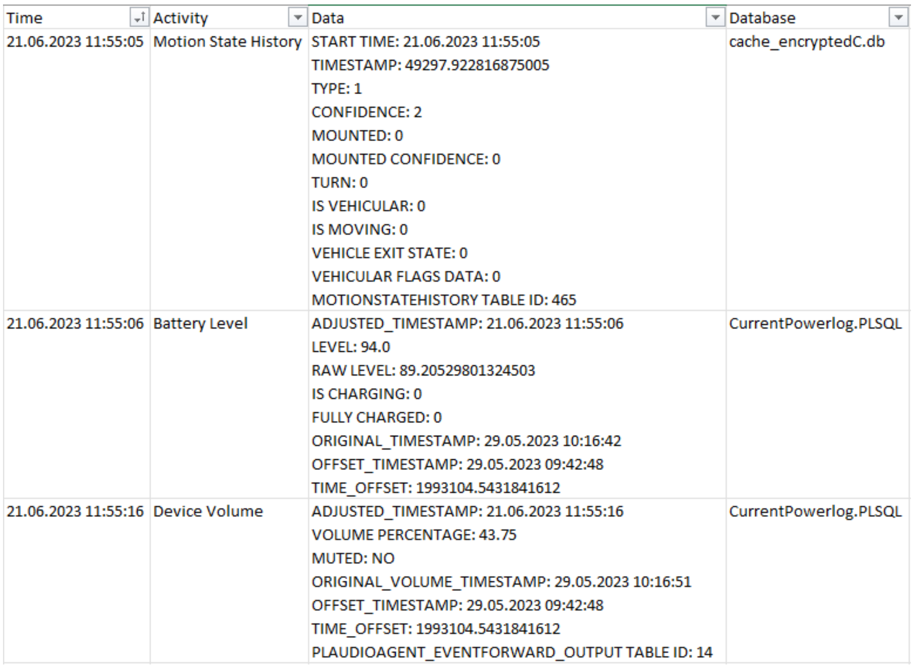
To add options for further improvements, a XML containing a definition for each activity within the results were created. This file allowed for exclusion of specific key-value pairs, in addition to specifying which timestamps to convert. In the case of this experiment, no key-value pairs were excluded from the final results. The makeup of the XML file is illustrated in Figure 9.



```xml
<?xml version="1.0" encoding="utf-8" ?>
<activities>
    <activity name="device volume">
        <field format="localtime">ADJUSTED_TIMESTAMP</field>
        <field>VOLUME PERCENTAGE</field>
        <field>MUTED</field>
        <field format="localtime">ORIGINAL_VOLUME_TIMESTAMP</field>
        <field format="localtime">OFFSET_TIMESTAMP</field>
        <field>TIME_OFFSET</field>
        <field>PLAUDIOAGENT_EVENTFORWARD_OUTPUT TABLE ID</field>
    </activity>
</activities>
```

Figure 9: XML structure for modifying properties in APOLLO

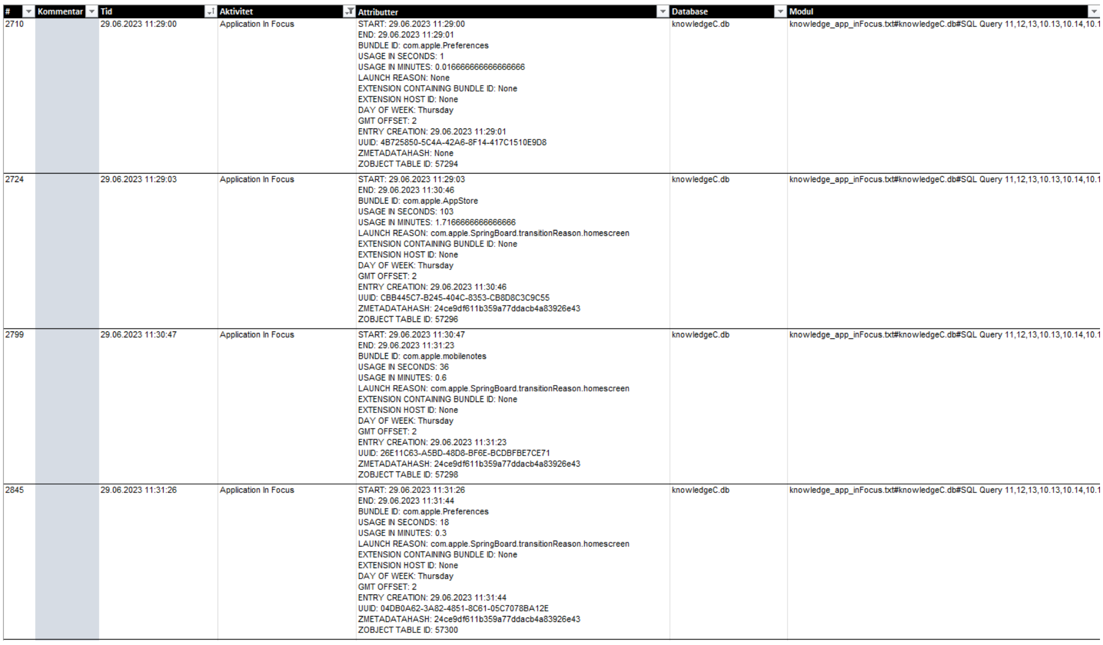
With the functionality mentioned in the above, the Python script converted the entirety of the APOLLO SQLite database to a data structure that harmonized with results exported from Cellebrite. The resulting output is illustrated in the Table 7.



Time	Activity	Data	Database
21.06.2023 11:55:05	Motion State History	START TIME: 21.06.2023 11:55:05 TIMESTAMP: 49297.922816875005 TYPE: 1 CONFIDENCE: 2 MOUNTED: 0 MOUNTED CONFIDENCE: 0 TURN: 0 IS VEHICULAR: 0 IS MOVING: 0 VEHICLE EXIT STATE: 0 VEHICULAR FLAGS DATA: 0 MOTIONSTATEHISTORY TABLE ID: 465	cache_encryptedC.db
21.06.2023 11:55:06	Battery Level	ADJUSTED_TIMESTAMP: 21.06.2023 11:55:06 LEVEL: 94.0 RAW LEVEL: 89.20529801324503 IS CHARGING: 0 FULLY CHARGED: 0 ORIGINAL_TIMESTAMP: 29.05.2023 10:16:42 OFFSET_TIMESTAMP: 29.05.2023 09:42:48 TIME_OFFSET: 1993104.5431841612	CurrentPowerlog.PLSQL
21.06.2023 11:55:16	Device Volume	ADJUSTED_TIMESTAMP: 21.06.2023 11:55:16 VOLUME PERCENTAGE: 43.75 MUTED: NO ORIGINAL_VOLUME_TIMESTAMP: 29.05.2023 10:16:51 OFFSET_TIMESTAMP: 29.05.2023 09:42:48 TIME_OFFSET: 1993104.5431841612 PLAUDIOAGENT_EVENTFORWARD_OUTPUT TABLE ID: 14	CurrentPowerlog.PLSQL

Table 7: APOLLO normalized output

Finally, the CSV file was imported into Excel and changed visually to align with the results from Cellebrite. This included an additional column for participant comments, and rearrangement of columns according to a standardized format. The final form of the data is illustrated Table 8.



#	Kommentar	Tid	Aktivitet	Attributter	Database	Modul
2710		29.06.2023 11:29:00	Application In Focus	START: 29.06.2023 11:29:00 END: 29.06.2023 11:29:01 BUNDLE ID: com.apple.Preferences USAGE IN SECONDS: 1 USAGE IN MINUTES: 0.016666666666666666 LAUNCH REASON: None EXTENSION CONTAINING BUNDLE ID: None EXTENSION HOST ID: None DAY OF WEEK: Thursday GMT OFFSET: 2 ENTRY CREATION: 29.06.2023 11:29:01 UUID: 4B725850-5C4A-42A6-8F14-417C1510E9D8 ZMETADATAHASH: None ZOBJECT TABLE ID: 57294	knowledgeC.db	knowledge_app_inFocus.txt#knowledgeC.db#SQL Query 11,12,13,10.13,10.14,10.1
2724		29.06.2023 11:29:03	Application In Focus	START: 29.06.2023 11:29:03 END: 29.06.2023 11:30:46 BUNDLE ID: com.apple.AppStore USAGE IN SECONDS: 103 USAGE IN MINUTES: 1.7166666666666666 LAUNCH REASON: com.apple.SpringBoard.transitionReason.homescreen EXTENSION CONTAINING BUNDLE ID: None EXTENSION HOST ID: None DAY OF WEEK: Thursday GMT OFFSET: 2 ENTRY CREATION: 29.06.2023 11:30:46 UUID: CBB445C7-B245-404C-8353-CB8D8C3C9C55 ZMETADATAHASH: 24ce9df611b359a77ddacb4a83926e43 ZOBJECT TABLE ID: 57296	knowledgeC.db	knowledge_app_inFocus.txt#knowledgeC.db#SQL Query 11,12,13,10.13,10.14,10.1
2799		29.06.2023 11:30:47	Application In Focus	START: 29.06.2023 11:30:47 END: 29.06.2023 11:31:23 BUNDLE ID: com.apple.mobilenotes USAGE IN SECONDS: 36 USAGE IN MINUTES: 0.6 LAUNCH REASON: com.apple.SpringBoard.transitionReason.homescreen EXTENSION CONTAINING BUNDLE ID: None EXTENSION HOST ID: None DAY OF WEEK: Thursday GMT OFFSET: 2 ENTRY CREATION: 29.06.2023 11:31:23 UUID: 26E11C63-A5BD-48D8-BF6E-BCDBFBE7CE71 ZMETADATAHASH: 24ce9df611b359a77ddacb4a83926e43 ZOBJECT TABLE ID: 57298	knowledgeC.db	knowledge_app_inFocus.txt#knowledgeC.db#SQL Query 11,12,13,10.13,10.14,10.1
2845		29.06.2023 11:31:26	Application In Focus	START: 29.06.2023 11:31:26 END: 29.06.2023 11:31:44 BUNDLE ID: com.apple.Preferences USAGE IN SECONDS: 18 USAGE IN MINUTES: 0.3 LAUNCH REASON: com.apple.SpringBoard.transitionReason.homescreen EXTENSION CONTAINING BUNDLE ID: None EXTENSION HOST ID: None DAY OF WEEK: Thursday GMT OFFSET: 2 ENTRY CREATION: 29.06.2023 11:31:44 UUID: 04DB0A62-3A82-4851-8C61-05C7078BA12E ZMETADATAHASH: 24ce9df611b359a77ddacb4a83926e43 ZOBJECT TABLE ID: 57300	knowledgeC.db	knowledge_app_inFocus.txt#knowledgeC.db#SQL Query 11,12,13,10.13,10.14,10.1

Table 8: APOLLO normalized output in its final form

3.12 Data analysis

As the research methodology is based on an explanatory sequential design, it included both quantitative and qualitative aspects. Therefore, the data analysis was segmented into two parts before being integrated into a whole.

Data collection:
The data collection phase includes responses from participants in two phases:

Firstly, participants were tasked to assess a series of information needs that were founded on the scenario. They were asked to tag all relevant artifacts and write short conclusions based on these findings. The main purpose of these conclusions was to gain an understanding of the rationale behind each participant's selection of artifacts, and align their understanding to the ground truth.

Secondly, respondents participated in semi-structured interviews to draw experiences from the preceding assignments. The intent behind the interview was to shed light on the participants' own experiences while solving assignments with each distinct tool. The interview followed a thematic approach to facilitate the identification of common themes across both groups.

Qualitative analysis:
The qualitative analysis was inspired by the thematic approach described by Braun and Clarke (2006). This approach involves a six step process whereby overarching themes can be extracted through a qualitative analysis:

1. Familiarization of data
2. Initial coding
3. Search for themes
4. Review of themes
5. Definition and naming of themes
6. Reporting

Inspired by the thematic approach by Braun and Clarke (2006), the qualitative analysis was conducted as follows:

The analysis process was initiated by a familiarization of the collected data in order to gain an understanding of the overall context. Distinct errors within each particular assignment were then identified based on the ground truth, including both misinterpretations and failures to assess set objectives. Throughout this analysis, common patterns that appeared across the assignments were identified. These patterns were then synthesized into broad themes, defined and given descriptive names. Errors identified throughout the assignments could then be attributed to concrete themes for further analysis.

A broad analysis and discussion of each synthesized theme were performed based on experiences gathered through review of the assignments (6.1). In addition, participant feedback gathered through the post-scenario interviews were interwoven with the themes to shed light on subjective aspects within the synthesized themes.

Quantitative analysis:
With errors categorized into broad themes, each error could be counted in a comparative analysis of the performance level of each group. Not only this, the placement of each error under distinct categories allowed for a measurement of their distribution across the themes. Note that the error rate is derived from the qualitative analysis and thematization, and is as such only an approximate estimation.

3.13 Biases

I have been practicing as a digital forensic investigator for four years at the time of finalizing this thesis. During this period, I have specialized in development of methodology for parsing and evaluating artifacts at activity level.

I have encountered numerous misconceptions during my practice, both regarding the meaning and reliability of artifacts. My awareness of such misconceptions does not only stem from my own misinterpretations, but from peer-reviewing reports written by others. I have not only attributed these misconceptions to inexperience, but also to the way in which digital forensic tools present their results. It is my personal experience with these challenges that has sparked the curiosity that shaped the theme of this thesis.

The research methodology borrows elements from my practical work experience, particularly concerning testing and verification of results. I possess an in-depth knowledge of both tools used in the study, and have therefore developed a set of pre-existing notions about their strengths and weaknesses. The research design was implemented with these notions in mind, especially in terms of not influencing the participant. In order to minimize my own influence over the participants, they were only provided with a consent form and an objective task description prior to the assignments. In addition, all participants conducted the assignments independently. Due to potential influence, the participants were not given any information about the themes of the post-scenario interviews, before finalizing the assignments.

The participants were however asked about their prior knowledge, familiarity, and preferences for various digital forensics tools in the post scenario interviews (5.1). In retrospect, these question might better be places prior to the assignments, as they may have been influenced by the preceding tasks.

3.14 Strengths and weaknesses

During the establishment of the research design, several strengths and weaknesses were identified. Awareness of strengths and weaknesses within a research design is vital when evaluating results, as it can offer valuable insight into potential biases both in the data collection and analysis of these results (Leedy, 2021, pp. 73–75).

Strengths:
The research design was created with a pragmatic approach, incorporating a close-to-realistic criminal incident to generate data on an iPhone. Each step of the scenario was documented as to establish a ground truth, something that underlay the repeatability of the quantitative aspects of the study.

Data from the iPhone was processed with actual digital forensics tools, and the participants in the study were digital forensic investigators practicing in law enforcement. These aspects are believed to increase the generalizability of the results, in terms of potential implications in real-world scenarios.

The study design implements a mixed-methods approach, collecting both quantitative and qualitative data. It bridges the gap between the tool and the human interpreter, making for a more holistic inquiry. Overall, the study design is believed to have relevancy for real-life practitioners, as a way to increase awareness towards possible misinterpretations and the need for quality assurance in digital forensics investigations.

Weaknesses:
The study employ a relatively small sample size, consistent of six respondents selected through a purposive sampling method. The sampling strategy started in an already known network, expanding the circle using the snowball sample method. The small sample size may have decreased the generalizability of the study, as it does not allow for meaningful statistical calculations of group variation and measurements of effect sizes. In addition, all respondents stem from a distinct environment, therefore, other environments could possibly yield variations in results based on factors such as education and work culture.

The study incorporates two different digital forensics analysis tools, Cellebrite Physical Analyzer and APOLLO. With regards to prior knowledge of the tools, the respondents were gravely more knowledgeable about the former tool, with all maintaining Cellebrite Certified Physical Analyst (CCPA) certifications (5.1). With regards to APOLLO, only some of the participants had prior knowledge, and none had any formal training. These factors could potentially skew the results in favor of the more well known tool.

4 Experiment

This chapter accounts for the data creation and collection phase of the study. The procedure employs the creation of a near-to-realistic scenario that mimics a criminal incident. Each significant event of the scenario was documented with timestamps, establishing a ground truth for evaluating participant responses. The purpose of the data creation phase was to generate data on an Apple iPhone that was paired with an Apple Watch. Data from the iPhone was acquired and processed with two distinct digital forensic analysis tools.

The respondents of the study were not given access to the ground truth, rather they were presented with an open background story, loosely based on events within the scenario. They were also presented with a series of information needs, tailored to events within the scenario, and asked to assess them with the available dataset. In order to measure differences in performance, the respondents were divided into two groups, with each group retrieving data from one of the two analysis tools. The groups assessments could then be evaluated, against the scenario's ground truth.

4.1 Data Creation

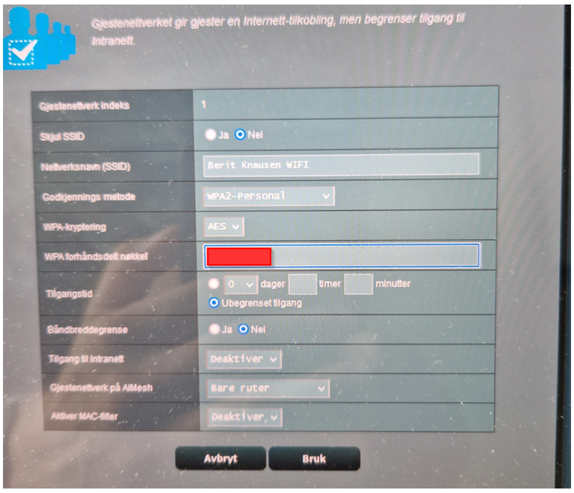
Two individuals played a part in the execution of the scenario: one person played the role of the victim, while another acted as a supervisor, guiding each action in accordance with a pre-arranged plan. As the scenario unfolded, both physical actions (such as walking, running, driving) and device interactions (playing music, browsing the web, sending messages) were documented in a timeline for future reference.

The actions of the scenario were kept to a moderate level, as to not generate an overwhelming amount of data for the latter assessments. After finalizing the scenario, data was acquired from the iPhone as detailed in Section 3.9.

4.1.1 External devices

One of the main objectives of the experiment was to generate data that could be associated with external devices, such as Wi-Fi routers, Bluetooth speakers, and vehicles. Not only would this allow data to be linked to identifiable devices, but would create data associated with physical locations.

To achieve this, a Wi-Fi network named "Berit Knausen WIFI" was set up at the victim's residence. The device was connected to the network using default settings, enabling it to auto-connect to the network in the future. Additionally, the device was paired with a Bluetooth speaker named "JBL Flip 5", also located within the residence. Connection and disconnections from these devices could thus be used to determine when the device was in proximity of the residence.



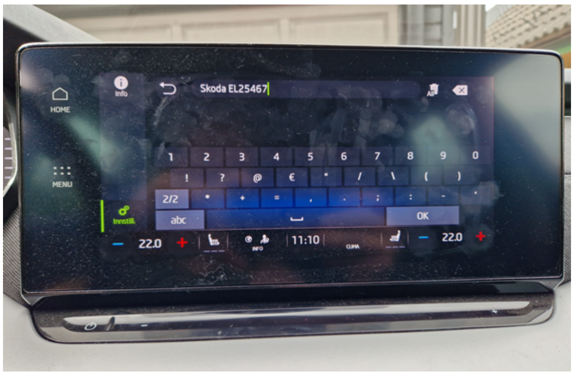
Gjestenettverket gir gjester en Internett-tilkobling, men begrenser tilgang til Intranett.

Gjestenettverk indeks	1
Skjul SSID	○ Ja ● Nei
Nettverksnavn (SSID)	Berit Knausen WIFI
Godkjennings metode	WPA2-Personal
WPA-kryptering	AES
WPA forhåndsdelt nøkkel	[redacted]
Tilgangstid	○ 0 dager timer minutter ● Ubegrenset tilgang
Båndbreddegrense	○ Ja ● Nei
Tilgang til Intranett	Deaktiver
Gjestenettverk på AiMesh	Bare ruter
Aktiver MAC-filter	Deaktiver

Figure 10: Setup of Wi-Fi network named Berit Knausen WIFI

Figure 11: Setup of Bluetooth speaker named JBL Flip 5

The scenario also included a vehicle with support for Apple Carplay. The Bluetooth and Wi-Fi names for vehicle were set as ”Skoda EL25467”, in order to mimic a license plate number. The iPhone was connected to CarPlay wireless, a process that utilized both the Bluetooth and Wi-Fi connection of the vehicle. As such, each connection to Apple CarPlay would be identified as ”Skoda EL25467” in the source data, linking the iPhone, not only to the vehicle but a fictitious license plate number.



Skoda EL25467

Figure 12: Setup of Apple CarPlay, named Skoda EL25467

4.1.2 How the scenario was created

The following section accounts for the main events of the scenario in broad strokes. For further details, a timeline of events is included in Section 4.1.3

Event 1: At the residence
The scenario was initiated at the victim's home, a two story house located in a residential area. While in the residence, the victim walked around the house listening to music with Spotify. To generate data that could be associated with the house, the device was connected to the Wi-Fi network, named "Berit Knausen WIFI", and later a Bluetooth speaker named "JBL Flip 5". An SMS message was received on the victim's iPhone, indicating that she was to be picked up. She sent a confirming SMS message, suggesting that she would be ready in about 10 minutes. Besides these initial SMS-messages, no further communications were present throughout the scenario.

Event 2: Connecting to Apple CarPlay
About 10 minutes preceding the SMS messages, the victim was picked up by the perpetrator, driving a vehicle whose Wi-Fi and Bluetooth names were set up as "Skoda EL25467". The victim walked outside, sat in the vehicle and connected her iPhone to Apple CarPlay. While in the vehicle, Apple CarPlay was used to play music with Spotify and for navigation with Apple Maps.

Event 3: Trip 1
The perpetrator drove the vehicle from the victim's residence to a health center. While traveling, the victim searched for "Indisk Restaurant" (Indian Restaurant) on her iPhone.

Event 4: Arrival at health center
The vehicle arrived at a parking lot in proximity to the health center. The victim walked from the car to the front entrance of the health center. She waited for about 5 minutes, before returning to the car, simulating a short errand at the center.

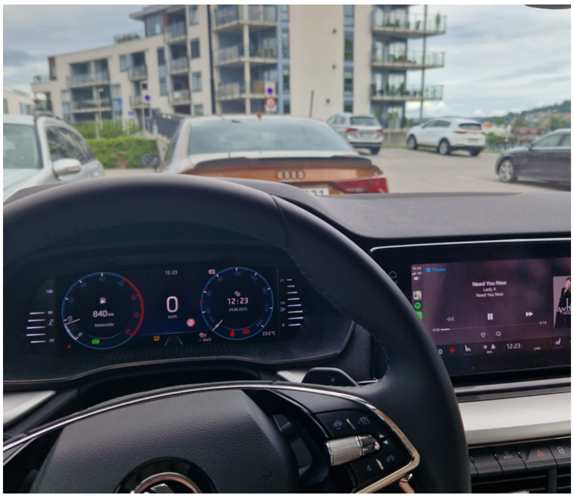


Figure 13: Scenario, at parking lot outside health center 12:24

Event 5: Trip 2
The perpetrator then drove for approximately 40 minutes towards a restaurant. While in the vehicle, the victim used Apple CarPlay to play music With Spotify while navigating to the destination with Apple Maps.

Figure 14: Scenario, entering restaurant at 13:15

Event 6: Eating at the restaurant
The vehicle was parked outside the destination. Both the victim and the perpetrator exited the vehicle and walked into the restaurant. They sat there for about 1 hour and 20 minutes, before returning to the vehicle.

Figure 15: Scenario, eating at restaurant 13:40

Event 7: Trip 3
After finishing eating, the victim and perpetrator returned to the vehicle in the parking lot.

Figure 16: Scenario, leaving restaurant at 14:41

Event 8: Trip 3
The vehicle was then driven back to the victim's residence, a trip that took about 1 hour. While in the vehicle, the victim used Apple CarPlay to play music while navigating with Apple Maps.

Event 9: Arrival at the residence
When approaching the residence, the device automatically connected to the residential Wi-Fi network "Berit Knausen WIFI." After a few minutes, the device also connected to the stationary Bluetooth speaker "JBL Flip 5". These connections were set up to indicate that the victim was back at her residence.

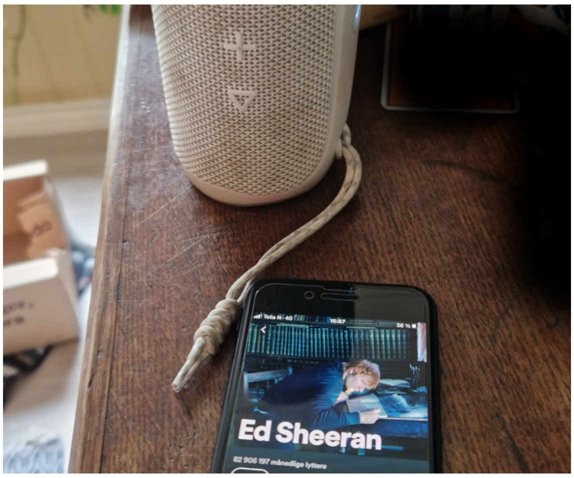


Figure 17: Scenario, playing music with Bluetooth speaker at 15:57

Event 10: Photo with generic camera functionality
The victim and the perpetrator remained at the residence for several hours. At one time, a photo was ”mistakenly” taken with the device’s camera, capturing a blurry photo, showing the legs of the perpetrator in the end of a sofa.

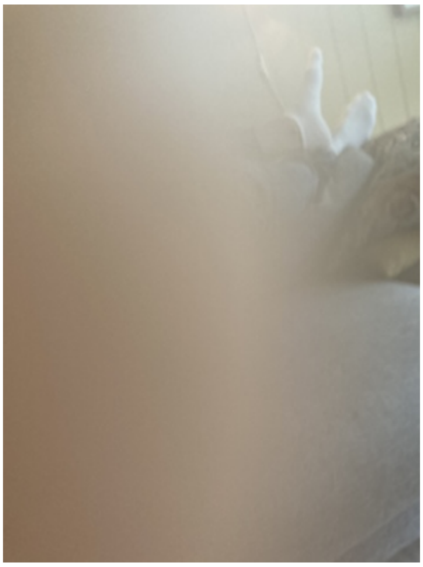


Figure 18: Scenario, photo 1 taken with the device at 18:07

Event 11: Connecting device to charger
About 30 minutes minutes later, the victim connected her iPhone to a charger and placed it on the floor.

Event 12: Opening Tinder application
Seconds after connecting the iPhone to the charger, the victim opened the dating application Tinder for a brief moment.

Event 13: Simulating physical conflict
A physical conflict was simulated by the victim falling to the floor. Jumping jacks were performed to increase the heart rate, simulating a physical confrontation. The heart rate increased quickly to about 156 beats per minute, before the Apple Watch was removed and placed on a flat surface. The Apple Watch was removed to indicate the point of death for the victim (In the fictitious scenario, the victim was still wearing the Apple Watch when discovered).

Event 14: Perpetrator escaping the crime scene, leaving the device on the ground
The perpetrator tore the victim's iPhone from its charger, suggesting that he was trying to dispose of the unit in haste. He fled the residence taking the iPhone with him, and stopped about 130 meters away. While handling the device, a photo of the ground was accidentally taken from a locked screen. The purpose of this was to simulate that the perpetrator was trying to gain access to the device's menus to activate flight mode. The perpetrator eventually managed to set the device in flight mode, before leaving the device on the ground.

Figure 19: Scenario, photo 2 taken with the device at 19:53

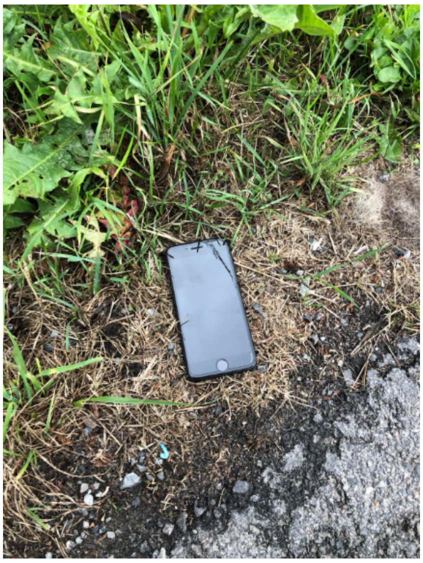


Figure 20: Scenario, photo of the device left on the ground, taken at 19:54

Event 15: Device found by the police
Several hours later, the device was found by the Police, seized and taken to a patrol car.

4.1.3 Detailed timeline

The table below provides a detailed account of the scenario as it unfolded, with each row signifying a single action or event. Each event was documented continuously on a separate device. This device used for documentation is not included in the scenario. The column ”Time from” specifies the start time of an event, while column ”Time to” specifies the end time where applicable. Column ”Description” contains a description of the event itself, while the last column ”Type” categorized the actions. The events described in the timeline could this serve as a ground truth when evaluating participant responses in the subsequent analysis.

Time from	Time to	Description	Type
11:25	11:59	Victim is at home and walks around the residence.	Health
11:28	-	Connected to the Wi-Fi network ”Berit Knausen WIFI.”	Wi-Fi
11:35	12:17	Plays music with Spotify.	Device
11:39	-	Connected to a Bluetooth speaker named ”JBL Flip 5.”	Bluetooth
11:43	-	Received an SMS: ”Hei Berit, hvordan ligger du an?”	Communications
11:45	-	Sent an SMS: ”Bare bra, er klar om 10 min. Gleder meg!”	Communications
11:50	-	Received an SMS: ”Utenfor nå!”	Communications
11:51	-	Sent an SMS: ”Kommer!”	Communications
11:55	-	Skoda arrives at the residence, and the victim walks out and sits in the car.	Health
11:57	-	Connected to Apple CarPlay (Skoda).	CarPlay
11:57	-	Connected to a Wi-Fi network named ”My Skoda 1396” (Skoda).	Wi-Fi
11:57	-	Connected to Bluetooth named ”Skoda EL25467” (Skoda).	Bluetooth
11:59	12:15	Skoda drives from the house to a health center.	Driving
12:09	12:11	Browses Safari for ”Indisk Restaurant.”	Device
12:18	12:22	Walks around the area, in front of the medical center, and back to the car.	Health
12:22	13:08	Plays music with Spotify.	Device
12:32	13:08	Skoda drives from the health center to the restaurant.	Driving
13:11	13:21	Walks from the parking lot into the restaurant.	Health
13:21	14:39	Stays at the restaurant.	Location
14:39	14:43	Walks from the restaurant back to the Skoda.	Health
14:43	15:41	Skoda drives from the restaurant back to the residence.	Driving
14:44	-	Connected to Apple CarPlay (Skoda).	CarPlay
14:44	-	Connected to a Wi-Fi network named ”My Skoda 1396” (Skoda).	Wi-Fi
14:44	-	Connected to Bluetooth named ”Skoda EL25467” (Skoda).	Bluetooth
15:43	-	Reconnected to the wireless network ”Berit Knausen WIFI.”	Wi-Fi
15:45	-	Walks from the car back to the residence.	Health
15:55	-	Reconnected to a Bluetooth speaker named ”JBL Flip 5.”	Bluetooth
18:07	-	Takes a photo with a Generic Camera, unclear photo with person in the background. The camera was started from the home screen.	Photo
19:41	-	Device plugged into charger.	Device
19:41	-	Victim opens Tinder application on the device	Device
19:46	-	Victim falls to the ground.	Health
19:47	-	Initiated jumping jacks to increase heart rate (simulated stress scenario).	Health
19:49	-	Heart rate reached 156 (simulated peak stress scenario).	Health
19:49	-	Removed Apple Watch (simulate death scenario).	Health
19:51	-	Device disconnected from charger.	Device
19:51	-	Phone is removed from the premises, walks about 130 meters, and leaves the iPhone on the ground.	Health

Time from	Time to	Description	Type
19:52	-	Lost connection to Bluetooth speaker ”JBL Flip 5.”	Bluetooth
19:53	-	Lost connection with Wi-Fi network ”Berit Knausen WIFI.”	Wi-Fi
19:53	-	Takes a photo with a Generic Camera of the ground. The camera was started from the locked screen.	Photo
20:00	-	Activated flight mode on the device.	Device
23:11	-	Device found by the Police, seized and taken to patrol car.	Health
-	-	No further actions performed	-

Table 9: Timeline for scenario

4.2 Background information given to the respondents

Preceding their assessment, respondents were given some background information based on the scenario. The purpose of this was to establish a context for the succeeding examination based upon the defined information needs. The following background information was presented for the participants:

> On June 29th, 2023, around 22:00, Berit Knausen was found dead in her home located on the outskirts of Skien. She was discovered by a friend who planned for a short visit to return some items. The friend explained that she had walked straight into the house as the front door was open upon her arrival.
>
> Berit was found lying face-up on the couch, and there were no visual signs of violence or break-in. According to her family, she had been in a good physical and mental health state for the last years, and nothing suggests she wanted to end her own life.
>
> Berit's iPhone was found on the side of a parking lot, approximately 130 meters from her house, and her Apple Watch was still on her wrist. Due to these circumstances, police have characterized her death as suspicious and want to investigate if Berit had been the victim of a crime.
>
> Forensic investigators have initiated a crime scene analysis, and police interviews of neighbors and others who may had recent contact with Berit are planned. Police have already seized Berit's mobile phone and started a preliminary content analysis.
>
> The lead investigator has compiled a list of information needs related to activities on Berit's mobile phone. As an investigator with competence in interpreting digital traces, you have been asked to assist in the assessment of these information needs.
>
> You have been provided a timeline of activities, exported to Excel, from police's mobile phone analysis tool. Due to the urgency of the situation, you are asked to answer each information need in a short and concise manner, and to make short comments on rows that contain data related to the tasks.

4.3 Information needs

A series of information needs were established based on the created scenario. The information needs were designed to cover a range of activities that could play significance in the initial stages of a real-life investigation. The information needs range from questions that typically can be answered by identifying a suitable activity, to questions that require detailed metadata from particular artifacts.

Participants were asked to provide short answers to the information needs, and to comment on rows that were the basis of this decision. The following information needs were presented to the participants:

Task 1: Analyze communications
The investigation is in its initial phases, and it is uncertain whether Berit had contact with anyone on the day of the incident. You are tasked to find out if Berit communicated with anyone prior to the incident.

Task 2: Connections to Wi-Fi networks
A wireless router has been found in the residence, and the lead investigator wants to know if the device was connected to the Wi-Fi network. Can you find any connections? If so, during which periods did these connections occur? Can you find the SSID and MAC address?

Task 3: Bluetooth connections A Bluetooth speaker has also been found in the residence. Has the device been connected to any Bluetooth devices? If so, what was it used for?

Task 4: Vehicle
One of Berit's friends has explained that she saw Berit in the passenger seat of a car on the same day that Berit died. Can you find any details to verify or falsify this claim?

Task 5: Browser history
You have been tasked to investigating the browser history on the device. Are there any activities that could provide context to Berit's actions prior to her death?

Task 6: Origin of photo 1
A photo with timestamp 18:07 has been found on Berit's iPhone. The image is unclear, but a person's legs appears to be in the background. The lead investigator would like you to investigate whether the photo was received or taken with the device. This could indicate whether she was with someone during the time before her died.

Task 7: Connection to charger A forensic investigator has noticed that the charging cable for Berit's iPhone is lying on the floor, approximately one meter away from the charging unit (which remains plugged in). This could be as possible situational trace, and the lead investigator has asked you to determine the last time the iPhone was plugged in.

Task 8: Last activities
The exact time of Berit's death is uncertain. You are tasked to identify any activities that could indicate her last signs of life. This includes her last interactions with the device, as well as health data of relevance.

Task 9: Airplane mode
When Berit's mobile phone arrived at the police station, it was already set to airplane mode. It is somewhat unclear whether Berit herself, the police, or someone else activated the setting. The lead investigator wants you to determine the last time the device was set to airplane mode.

4.4 Interview

Following the analysis tasks, the participants were asked to participate in an interview. The purpose of this interview was to gain insight into potential challenges that occurred during the assessments. This data was collected in order to gain background information and experience from the two groups.

The participants were asked the following questions:

Question 1: Forensic tool knowledge

Which forensic mobile phone analysis tools are you knowledgeable about? Which tool do you prefer for pattern of life analysis? Why do you prefer this specific tool?

Question 2: Analysis of data at activity level
What experience do you have in analysis of data at activity level?

Question 3: Technical challenges
Did you encounter any technical challenges while addressing the information needs?

Question 4: Missing data
Was any information difficult to find, or missing from the dataset? What were the challenges in finding this information?

Question 5: Interpretation
Was there information you were unable to interpret or understand? What were these challenges? Did you encounter too few or too many details?

Question 6: Accuracy and reliability
Did you consider the accuracy and reliability of the data you interpreted? Was there any information that you felt needed to be double checked or confirmed with another method?

Question 7: Tool selection
Do you have any thoughts on whether using different tools to address the information needs would have influenced your findings? If so, in what way?

Question 8: Competency
Were there any special skills or knowledge you needed to interpret and analyze the data?

Question 9: Additional comments
Is there anything you would like to add or comment?

5 Results and analysis

The purpose the chapter is to analyze and discuss results from the groups' assessments. It seeks to determine if there were significant differences between the groups and whether the variance can be linked to factors within each tools' presentation layer.

The analysis starts with an overview of the participants' prior knowledge, familiarity, and preferences for various digital forensics tools. Although this data were collected in the post-scenario interviews, it is a natural starting point as provides a broad outlook on the sample group. In retrospect, it is worth noting that questions regarding the participants' prior knowledge might have been better placed beforehand, as their answers may have been influenced by the assessments.

It continues by analyzing the participants' responses to the information needs and how well the answers align to the ground truth. Potential errors are analyzed to determine if they stem from factors within each tool's presentation layers.

The remaining parts of the post-scenario interview are presented and analyzed to shed light on the participants subjective experiences during the assignments.

5.1 Participants background and experience

This section explores the participants familiarity and preferences for various digital forensic tools, their educational background, work experience, and their prior experience working with data at activity level.

5.1.1 Background and education

All of the participants have relevant education and practical experience within digital forensics. Three of the participants have master's degrees, two have bachelor's degrees, and one participant has multiple years of experience and extensive training within digital forensics. The respondents work experience ranges from 1-7 years, with backgrounds in both the private sector and law enforcement.

5.1.2 Familiarity with forensic tools

The respondents were asked to describe their familiarity, preferences, and experience with digital forensic tools:

The respondents reported familiarity with a range of proprietary forensic tools, including Cellebrite Physical Analyzer, Magnet Axiom, and MSAB XRY. Three of the participants were familiar with open source forensics tools including APOLLO and iLEAPP. Alongside this, they had experience with tools suitable for manual investigation of specific file formats, such as SQLite databases and information property files.

All of the participants highlighted Cellebrite Physical Analyzer as their primary tool during case work. When asked about the tool's capabilities, qualities such as user-friendliness and clear presentation of data were highlighted. When asked further about the rationale behind this tool choice, the participants expressed that they were more influenced by factors such as familiarity, training, and the tool's widespread recognition, than the tool's inherent qualities.

The majority of participants did not depend on a single tool alone, but prefered a combination of tools. A common strategy was to initiate a triage with Cellebrite Physical Analyzer, then merge the analysis with more specialized tools at need. These respondents emphasized that the approach supports the concept of dual-tool verification, which implies the use of at least two different tools to cross-verify results. Some of the participants noted differences in results when comparing the most common proprietary tools, indicating that by using several tools, more artifacts could be uncovered. This multi-tool approach was particularly emphasized by two of the participants, who expressed general skepticism about presenting data from one tool alone.

The participants who had prior experience with APOLLO highlighted the tool's ability to parse more artifacts with a greater level of detail, compared to that of Cellebrite Physical Analyzer. None of these participants have any formal training with the tool.

All of the participants in the study have formal training with Cellebrite Physical Analyzer, including at least one certification. Even though the majority of the respondents displayed a preference for Cellebrite Physical Analyzer in their casework, the rationale behind this choice seems to be based on the tool's widespread acceptance rather than its inherent qualities. Interestingly, most of the participants utilize a multi-tool approach, starting with Cellebrite Physical Analyzer and transitioning to specialized tools when needed. Despite the lack of any formal training, three of the participants highlighted APOLLO as a tool that presents more artifacts with greater detail compared to Cellebrite.

5.1.3 Experience with data at activity level

The participants were asked to describe their knowledge and prior experience with analysis of data at the activity level. The respondents had varying experiences, with one participant claiming extensive knowledge, three reporting moderate experience, and two stating limited experience.

The participant who claimed extensive knowledge, reported familiarity with most well-known proprietary forensic tools. Despite this, they preferred open-source tools such as APOLLO and iLEAPP when engaging in advanced forms of analysis. They emphasized the importance of verifying results through extensive testing, including post-test comparisons between test results and findings. They regarded this methodology as a necessary precursor when drawing conclusion, especially due to regular software updates. They added that due to little to no documentation from the vendors, the creation of a sound knowledge base for interpretation is a full-time endeavor by its own.

One of the participants who claimed moderate knowledge, reported to have some professional experience with analysis of activity data with Cellebrite. Even though they had not used APOLLO in a professional setting, they preferred the tool to Cellebrite due to its granularity of artifacts.

The second respondent with moderate knowledge, had performed some analysis of activities through their work, but only at a basic level. Despite claiming moderate experience, both of these participants had undergone training in timeline analysis and incident response through their education.

The third participant with moderate experience, had not performed any such analysis tasks themselves, but had gained a lot of insight by peer reviews and relevant literature.

None of the participants who reported limited experience had conducted such analysis on iOS devices. One of these participants had extensive experience with timeline and log file analysis from a professional context, while the other participant had gained some experience through their education.

Despite varying experience, all of the participants demonstrated an understanding of the concept of pattern of life analysis and its practical use cases.

5.2 Interpretation of results

The first section compares and analyzes error rates between the APOLLO and Cellebrite group in their assessments of the various information needs. The aim is not merely to showcase the groups error rates for each particular assessment, but to identify the underlying factors that contributed to the failures. By analyzing each assignment individually, an insight into common factors in each tool's presentation layer can be identified. Results from this analysis are synthesized and discussed further in Section 6.1.

5.2.1 Task 1: Communications

> ”The investigation is in its initial phases, and it is uncertain whether Berit had contact with anyone on the day of the incident. You are tasked to find out if Berit communicated with anyone prior to the incident.”

In the first assignment (4.3), the participants were asked if they could identify whether the victim had communicated with anyone prior to her death. The primary objective was to assess if the groups differed while solving basic tasks, and secondly, to what degree participants utilized additional information such as application usage and notifications to support the communications.

Failure to utilize additional contextual information could in come instances lead to erroneous conclusions as several communications applications synchronize data across connected devices. As an example, iMessage supports synchronization of both messages and call history across iCloud-connected devices.[7]

APOLLO:
Two out of three participants in the APOLLO group identified all SMS messages in accordance with the ground truth (4.1.3). They also included SMS read confirmations to underpin that the messages in fact had been read on the device.

One of the participants included a secondary source of metadata in the form of Contact interactions. The significance of this artifact is its potential to showcase information about deleted communications, an important factor in many investigations. Interestingly, the most experienced of the participants only selected SMS read confirmations, missing out on two outgoing messages.

Considering that iMessage allows for synchronization of call logs and messages across iCloud-connected devices, the participants could with benefit have included application usage to underpin that the messages were written on the device. While the group demonstrated some uniformity by highlighting SMS messages and read confirmations, their strategy seemed more aligned with a search for activities with suitable names, rather than a systematic approach based on prior knowledge and experience.

Cellebrite:
As with the first group, only two of three participants discovered all of the SMS messages. The two participants who found all of the messages did not provide any additional activities to underpin the communications.

The participants who failed to provide all messages used a combination of messages and application usage to demonstrate that the SMS application had been opened in relevant time frames. The addition of application usage enabled the participant to identify that Tinder had been open later in the evening, suggesting that a manual inquiry into its database could provide value for the investigation. The failure to include all of the SMS messages appears to be related to neglect rather than competency.

General findings:
Even though the groups were introduced with a basic task, only two out of three participants in each group succeeded in finding all relevant communications.

The reasoning behind these failures appears to differ between the groups. While the participant in the Cellebrite group overlooked a single message, the participant in the APOLLO group misinterpreted the activity, ”SMS Chat - Message Read”, believing that it contained all SMS communications. This misunderstanding might rest on the fact that APOLLO presented SMS-related information into three discrete activities, rather than in one merged view as the case with Cellebrite.

The implications of missing some or all of these SMS messages in a real-world context could undoubtedly affect the investigation. The messages act as a precursor to all proceeding action and establishes a direct link to the perpetrator.

[7] https://support.apple.com/guide/icloud/set-up-messages-mm0de0d4528d/icloud

Summary of errors:

- APOLLO: One participant misinterpreted an activity name due to naming conventions
- Cellebrite: One participant missed a message due to neglect

5.2.2 Task 2: Connections to Wi-Fi networks

> "A wireless router has been found in the residence, and the lead investigator wants to know if the device was connected to the Wi-Fi network. Can you find any connections? If so, during which periods did these connections occur? Can you find the SSID and MAC address?"

The second task (4.3) required the respondents to identify time frames for Wi-Fi connections with associated SSID names and BSSID addresses. If the participants could identify these connections, they would have the necessary information to tie the device to the victim's residence at given time frames. Additionally, it would aid them in Task 4 (5.2.4), as it would help them determine that the victim's phone was connected to a vehicle. As last it would give valuable information in Task 8 (5.2.8), by pinpointing the moment the device was removed from the residency by the perpetrator.

APOLLO:
All respondents in the APOLLO group successfully identified the name of the victim's home Wi-Fi network, and were able to establish connection intervals in accordance with the ground truth (4.1.3).

APOLLO parses network-related information from two distinct databases, KnowledgeC.db, and Currentpowerlog.PLSQL into a single activity named "WiFi connection". While the participants succeeded in using information from the former database they failed to consider information parsed from the latter. While the database KnowledgeC.db contains information about Wi-Fi SSID's and connection time frames for distinct connections, Currentpowerlog.PLSQL provides granular information about Wi-Fi channel usage with corresponding time stamps. By overlooking the latter source, they missed the opportunity to showcase even more precise timestamps regarding Wi-Fi connections and disconnections.

The assignment also required the participants to identify the BSSID address of the victim's home network. While all respondents reported that they discovered potential BSSIDs, they refrained from tagging them due to uncertainties. In the post-scenario interview, it was revealed that two participants attempted to convert BSSIDs from APOLLO's activity "WIFI Location". Despite these efforts, information contained within this particular activity originates from the database `cache_encryptedB.db`, containing harvested Wi-Fi locations unrelated to the information need (Whiffin, 2022). The challenges of converting these MAC addresses are understandable, as they are stored in an base 10 numeric format. The last participant also made a mistake by attempting to convert a UUID in the activity "WiFi connection", mistaking it for a BSSID address.

Cellebrite:
Contrary to the APOLLO group, none of the participants in the Cellebrite group were able to identify Wi-Fi connections that corresponded with the ground truth (4.1.3). One participant refrained entirely from responding to the assessment, while the other participants both selected 29 harvested Wi-Fi locations.

The list of networks selected by the participants resided within a category named Wireless Networks. Although the name suggests actual Wi-Fi connections, information within the source file information column reveals that the information is parsed from the database `cache_encryptedB.db`, known to consist of harvested Wi-Fi locations (Whiffin, 2022). While not obvious, additional information within the source file columns gives hints of its purpose, as it references a table named WifiLocation.

By delving into the tool's treatment of information from this database, it becomes clear that information from `cache_encryptedB.db` is segregated into two distinct categories, Wireless Network: Location, and Wireless Networks. Interestingly, the only difference between these two categories is that the former contains harvested Wi-Fi networks with coordinates, while the latter contains harvested Wi-Fi networks that lack coordinates. Based on this inquiry, it becomes apparent why the participants misinterpreted the harvested Wi-Fi networks as actual Wi-Fi connections. Cellebrite's somewhat ambiguous distinction between networks with and without location data can easily be misunderstood by both untrained and trained practitioners. Interestingly, one of the participants that tagged these networks, warned about these specific cases and how they could be misinterpreted.

However, the most concerning aspect of the assessment was that none of the participants were able to identify the activity that actually showcases Wi-Fi connection intervals. Surprisingly, this activity does not reside within the categories "Wireless Networks" or "Device Connectivity", but within the category "Device Events". Cellebrite presents these connections with the ambiguous descriptors "Connect" and "Disconnect", which arguably are not easily recognizable. Despite that this activity provides the correct connection intervals, no information regarding the SSID name or BSSID addresses are provided.

General findings:
The primary objective of the assignment was to identify Wi-Fi connection intervals with their associated SSID names and BSSID addresses. Firstly, the SSID names were required to distinguish between connections to different access points and to link the victim's phone to specific devices. Secondly the connection intervals were essential to showcase the devices movements, both while residing within the residence and while connected to the vehicle. While not essential, the addition of a BSSID address would further tie the phone to specific devices.

Despite that the APOLLO group succeeded in identifying the victims home network SSID, with the correct connection intervals, they encountered challenges while attempting to find the BSSID of the network. Their attempts to decode erroneous BSSIDs addresses may suggest that there is room for improvements in terms of training and the tool's naming conventions.

While the APOLLO group largely succeeded in the assignment, the Cellebrite group struggled considerably. The group was neither able to identify the home network, nor any connection intervals of substance. The reasoning behind the failures appears to be tied to Cellebrite's ambiguous naming conventions and lack of contextual information. Despite that the dataset contained information about the correct Wi-Fi connection intervals, none of the participants were able to recognize them.

Summary of errors:

- APOLLO: Three participants failed to identify the BSSID of the victim's home Wi-Fi network due to lacking details
- Cellebrite: Three participants failed to identify the correct activity due to naming conventions
- Cellebrite: Two participants selected an erroneous activity due to naming conventions
- Cellebrite: Three participants failed to identify the Wi-Fi SSID due to naming conventions and lack of details
- Cellebrite: Three participants failed to identify the Wi-Fi BSSID due to naming conventions and lack of details

5.2.3 Task 3: Bluetooth connections

> "A Bluetooth speaker has also been found in the residence. Has the device been connected to any Bluetooth devices? If so, what was it used for?"

In the third task (4.3), respondents were informed that a Bluetooth speaker was found in the victim's residence, and asked if the victim's phone had connected to it. Is do, they were to provide details regarding its usage.

Similar to Task two (5.2.2), information about connections to the Bluetooth speaker had the potential to associate the device with the victim's residence at specific time frames. In addition, it could provide valuable information about the victim's movements between her residence and the vehicle in Task four (5.2.4). Lastly, it could provide value in Task eight (5.2.8), as loss of connection to the speaker corresponds with the time the device was removed from the residence by the perpetrator.

By including an additional task related to wireless connections, the participants had two potential sources of information, suitable for determining movements of the device. However, to clearly distinguish one Bluetooth connection from another, identifiable information such as Bluetooth device names and MAC addresses was required. Additional information regarding the use of the speaker, could also provide context to the victim's behavior on the day of the incident.

APOLLO:
All respondents in the APOLLO group successfully established that the victim's phone had been connected to the Bluetooth speaker. They provided details about the device's name and MAC address, and established connection intervals that aligned with the ground truth (4.1.3). Despite that the phone had been connected to various Bluetooth devices, they were able to single out the Bluetooth speaker and exclude irrelevant connections.

Only one of the participants fulfilled the second requirement of the task, namely to provide information regarding its usage. This participant was able to showcase its usage through the "Now Playing" activity, demonstrating the exact music being played on Spotify.

Cellebrite:
Despite that all of the participants in the Cellebrite group were able to identify Bluetooth connections, they were unable to distinguish between connections to different devices. As a consequence, they selected connections to both relevant and non-relevant devices. The reason behind their failures seems to be tied to the tool's lack of identifiable information, such as Bluetooth names and MAC addresses.

A common strategy between the respondents, was to select the activities "BluetoothA2DPOutput" and "Speaker" both containing information parsed from the database KnowledgeC.db, a database that inherently contains both Bluetooth names and MAC addresses.

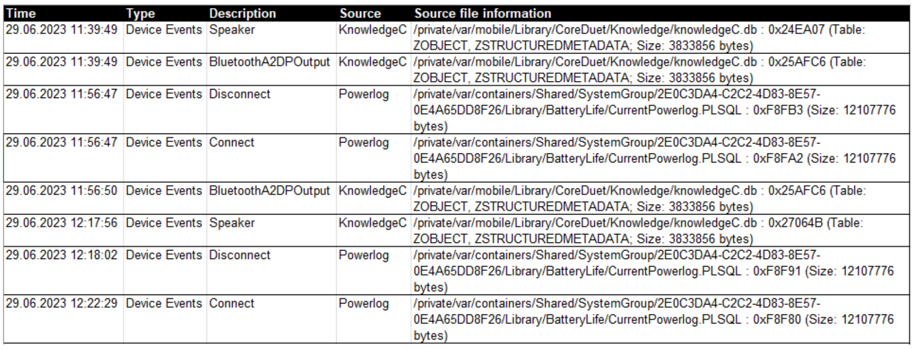
Despite failing to produce information about its usage one, participant was able to showcase the use of Spotify through the tool's application usage logs, inferring that Spotify was being used in the same time span as some of the Bluetooth connections.

One participant mistakenly interpreted Wi-Fi connections for connections to Bluetooth devices. As already discussed in Task two (5.2.2), the selected Wi-Fi activity is identified by its somewhat ambigious descriptors "Connect" and "Disconnect".

General findings:
All respondents in the APOLLO group successfully identified the name and MAC address of the Bluetooth speaker. They also demonstrated the connection intervals that aligned with the ground truth (section 4.1.3). In addition, one respondent identified that Spotify was in use, and were able to give specifics regarding the exact tracks being played.

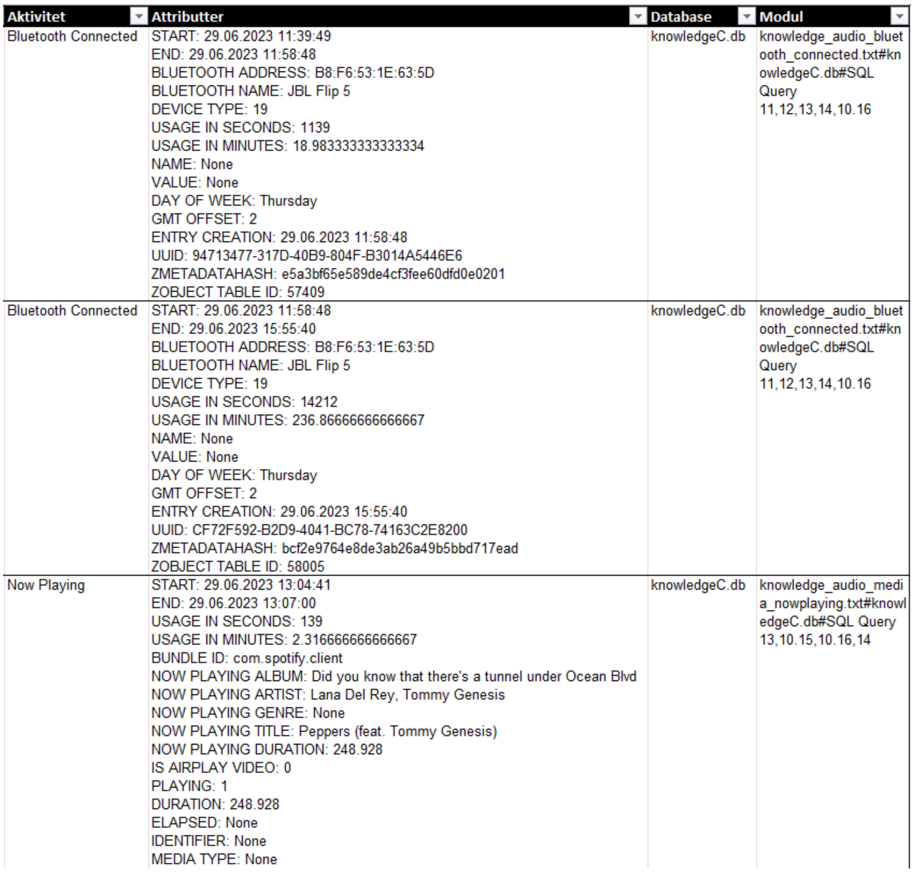
In contrast, participants in the Cellebrite group struggled to distinguish between connections to different Bluetooth devices. As a result they selected of both relevant and non-relevant Bluetooth devices.

The richness of data provided by APOLLO, such as time frames, Bluetooth name and MAC addresses, gave the group a significant advantage over Cellebrite. This advantage is demonstrated by Table 10 and 11, where it is evident why the Cellebrite group struggled with the assessment. Cellebrite's lack of details not only hindered the group from distinguishing connections from each other, but a compounding effect was seen, as a participant misinterpreted Wi-Fi connections for Bluetooth.



Time	Type	Description	Source	Source file information
29.06.2023 11:39:49	Device Events	Speaker	KnowledgeC	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x24EA07 (Table: ZOBJECT, ZSTRUCTUREDMETADATA; Size: 3833856 bytes)
29.06.2023 11:39:49	Device Events	BluetoothA2DPOutput	KnowledgeC	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x25AFC6 (Table: ZOBJECT, ZSTRUCTUREDMETADATA; Size: 3833856 bytes)
29.06.2023 11:56:47	Device Events	Disconnect	Powerlog	/private/var/containers/Shared/SystemGroup/2E0C3DA4-C2C2-4D83-8E57-0E4A65DD8F26/Library/BatteryLife/CurrentPowerlog.PLSQL : 0xF8FB3 (Size: 12107776 bytes)
29.06.2023 11:56:47	Device Events	Connect	Powerlog	/private/var/containers/Shared/SystemGroup/2E0C3DA4-C2C2-4D83-8E57-0E4A65DD8F26/Library/BatteryLife/CurrentPowerlog.PLSQL : 0xF8FA2 (Size: 12107776 bytes)
29.06.2023 11:56:50	Device Events	BluetoothA2DPOutput	KnowledgeC	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x25AFC6 (Table: ZOBJECT, ZSTRUCTUREDMETADATA; Size: 3833856 bytes)
29.06.2023 12:17:56	Device Events	Speaker	KnowledgeC	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x27064B (Table: ZOBJECT, ZSTRUCTUREDMETADATA; Size: 3833856 bytes)
29.06.2023 12:18:02	Device Events	Disconnect	Powerlog	/private/var/containers/Shared/SystemGroup/2E0C3DA4-C2C2-4D83-8E57-0E4A65DD8F26/Library/BatteryLife/CurrentPowerlog.PLSQL : 0xF8F91 (Size: 12107776 bytes)
29.06.2023 12:22:29	Device Events	Connect	Powerlog	/private/var/containers/Shared/SystemGroup/2E0C3DA4-C2C2-4D83-8E57-0E4A65DD8F26/Library/BatteryLife/CurrentPowerlog.PLSQL : 0xF8F80 (Size: 12107776 bytes)

Table 10: Detail level of activities in Cellebrite



Aktivitet	Attributter	Database	Modul
Bluetooth Connected	START: 29.06.2023 11:39:49 END: 29.06.2023 11:58:48 BLUETOOTH ADDRESS: B8:F6:53:1E:63:5D BLUETOOTH NAME: JBL Flip 5 DEVICE TYPE: 19 USAGE IN SECONDS: 1139 USAGE IN MINUTES: 18.983333333333334 NAME: None VALUE: None DAY OF WEEK: Thursday GMT OFFSET: 2 ENTRY CREATION: 29.06.2023 11:58:48 UUID: 94713477-317D-40B9-804F-B3014A5446E6 ZMETADATAHASH: e5a3bf65e589de4cf3fee60dfd0e0201 ZOBJECT TABLE ID: 57409	knowledgeC.db	knowledge_audio_bluetooth_connected.txt#knowledgeC.db#SQL Query 11,12,13,14,10.16
Bluetooth Connected	START: 29.06.2023 11:58:48 END: 29.06.2023 15:55:40 BLUETOOTH ADDRESS: B8:F6:53:1E:63:5D BLUETOOTH NAME: JBL Flip 5 DEVICE TYPE: 19 USAGE IN SECONDS: 14212 USAGE IN MINUTES: 236.86666666666667 NAME: None VALUE: None DAY OF WEEK: Thursday GMT OFFSET: 2 ENTRY CREATION: 29.06.2023 15:55:40 UUID: CF72F592-B2D9-4041-BC78-74163C2E8200 ZMETADATAHASH: bcf2e9764e8de3ab26a49b5bbd717ead ZOBJECT TABLE ID: 58005	knowledgeC.db	knowledge_audio_bluetooth_connected.txt#knowledgeC.db#SQL Query 11,12,13,14,10.16
Now Playing	START: 29.06.2023 13:04:41 END: 29.06.2023 13:07:00 USAGE IN SECONDS: 139 USAGE IN MINUTES: 2.316666666666667 BUNDLE ID: com.spotify.client NOW PLAYING ALBUM: Did you know that there's a tunnel under Ocean Blvd NOW PLAYING ARTIST: Lana Del Rey, Tommy Genesis NOW PLAYING GENRE: None NOW PLAYING TITLE: Peppers (feat. Tommy Genesis) NOW PLAYING DURATION: 248.928 IS AIRPLAY VIDEO: 0 PLAYING: 1 DURATION: 248.928 ELAPSED: None IDENTIFIER: None MEDIA TYPE: None	knowledgeC.db	knowledge_audio_media_nowplaying.txt#knowledgeC.db#SQL Query 13,10.15,10.16,14

Table 11: Detail level of activities in APOLLO

Summary of errors:

- Cellebrite: Three participants failed to identify the Bluetooth device name due to lack of details
- Cellebrite: Three participants failed to identify the Bluetooth MAC address due to lack of details
- Cellebrite: Three participants failed to differentiate between Bluetooth devices due to lack of details
- Cellebrite: One participants selected an erroneous activity due to naming conventions

5.2.4 Task 4: Car Play

> ”One of Berit’s friends has explained that she saw Berit in the passenger seat of a car on the same day that Berit died. Can you find any details to verify or falsify this claim?”

In this assignment, the respondents were informed that a friend of the victim believed she had seen Berit in the passenger seat of a vehicle, on the same day as the incident. Thus, they were asked to identify artifacts that could either support or refute this claim.

To prepare for the scenario, the Bluetooth name and Wi-Fi SSID of the vehicle were changed to mimic a license plate number belonging to a Skoda. This setup was designed to provided actionable information, as it establishes a direct link between the device and the perpetrator’s vehicle.

APOLLO:
All participants in the APOLLO group identified connections to the vehicles Apple CarPlay in accordance with the ground truth (4.1.3). They were able to provide information about the correct connection intervals, name, and at lest one identifier for the vehicle.

One participant relied solely on Bluetooth connections, while a second included both Bluetooth and Wi-Fi connections. The last participant used Bluetooth, Wi-Fi and ”Vehicle Park History”, producing a total of three artifacts that could be linked to the vehicle.

When commenting on the task, all respondents concluded that the victim had been in a vehicle, suggesting a Skoda with the license plate number EL25467. While not specifically required by the task, one participant also included additional contextual information, showcasing that the victim had used navigation software while playing music with Spotify through CarPlay.

Cellebrite:
The Cellebrite group neither relied on activities related to Bluetooth connections or Wi-Fi in the assessment. The exclusion of these activities seems related to the group’s failures in the two preceding tasks (5.2.2 and 5.2.3).

The participants did find an activity with the descriptor ”CarAudio”, indicating that the victim’s phone had been connected to a vehicle. While the artifact did provide time frames that aligned with the ground truth, it neither contained the name of the vehicle nor any other form of identifiable information. Some of the participants included additional contextual information, such as application and data usage, suggesting the use of Spotify and navigation software.

All participants expressed frustration due to lack of details: One participant remarked that they found very few details, leading them to guesswork with regards to the meaning of the artifacts. Another participant mentioned that although the artifact was most probably related to CarPlay, this could not be established with certainty due to a lack of details. The last participant shared a similar view, and remarked that the data lacked information related to any specific vehicles.

General findings:
The APOLLO group found a total of three source of information, Bluetooth, Wi-Fi, and ”Vehicle Park History”, all containing information that could be linked to a specific vehicle.

While the Cellebrite group found indications of CarPlay usage, neither participant was able to find any identifiable information. Due to a confounding effect based on the two previous assignments, the group failed to consider information about Bluetooth and Wi-Fi usage:

1. During the assessment of Wi-Fi activities in Task 2 (5.2.2), none of the respondents were able to find the correct Wi-Fi activity. Instead, they relied on crowd sourced Wi-Fi locations that had little relevance to the assignment.

2. Due to lack of details regarding Bluetooth connections in Task 3 (5.2.3), the participants were unable to differ between connections to various Bluetooth devices. In their results, participants included connections that spanned the entirety of the dataset, containing both connection to the home speaker and the vehicle.

This specific task was designed to contain actionable intelligence in a real life setting. By including identifiable information, such as the license plate number, a direct link could be made between the victim's phone and the perpetrator's vehicle. The differences in outcome is therefore concerning, as lack of finding such details could have impaired the initial stages of an investigation.

Summary of errors:

- Cellebrite: Three participants failed to identify the Bluetooth device name of the Vehicle due to lack of details
- Cellebrite: Three participants failed to identify the Bluetooth MAC address of the Vehicle due to lack of details
- Cellebrite: Three participants failed to identify the Wi-Fi SSID of the Vehicle due to lack of details
- Cellebrite: Three participants failed to identify the Wi-Fi BSSID of the Vehicle due to lack of details
- Cellebrite: Three participants failed to differentiate between Bluetooth devices due to lack of details

5.2.5 Task 5: Browser history

> "You have been tasked to investigate the browser history on the device. Are there any activities that could provide context to Berit's actions prior to her death?"

In the fifth task, the respondents were asked if they could provide any browser history for the victim's phone. A limited amount of web searches was performed on the device, one for the current weather, and one for the term "indisk restaurant" (Indian restaurant). The aim was produce some context regarding the victim's movements.

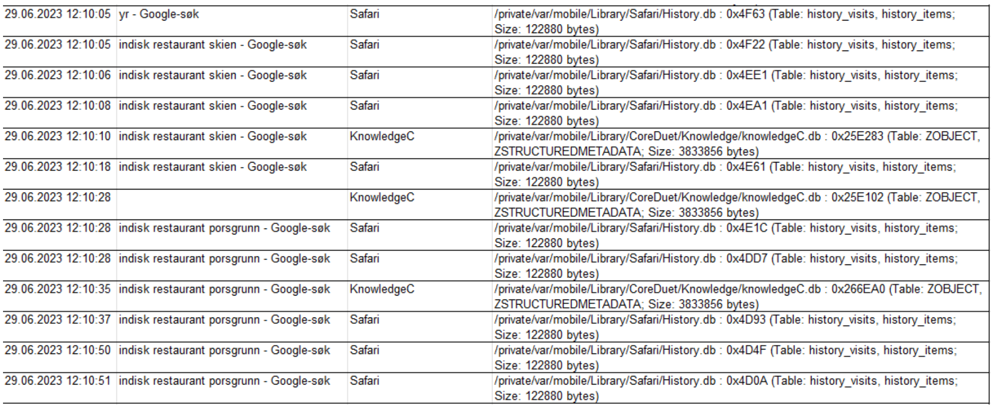
APOLLO:
All participants in the APOLLO group identified relevant browsing history. Common for the respondents were use of the same two activities, parsing browser history from History.db and knowledgeC.db.

Cellebrite:
The Cellebrite group also identified relevant browsing history through two activities, "Web History" and "Searched items". One participant supplemented the browser history with activities that demonstrated use of the Safari browser.

General findings:
None of the participants experienced issues while identifying the relevant browsing history.

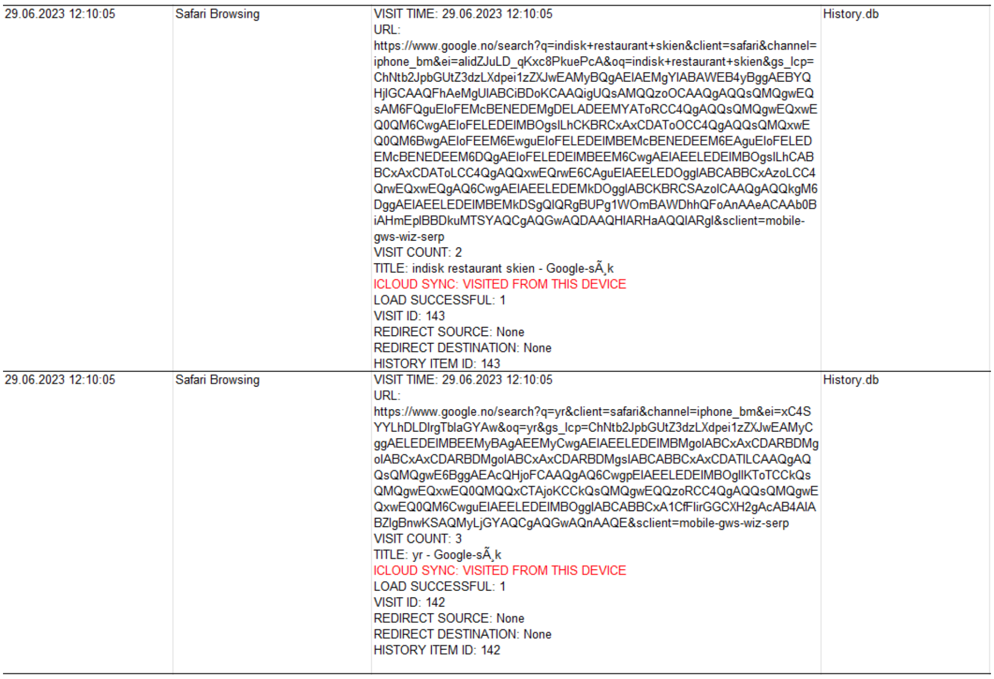
Although the following was not included as an aspect of the scenario, the tools varied regarding the detail level they provided for the phones browsing history. The Safari browser supports synchronization of web history across iCloud-connected devices[8]. One area of concern is therefore that Cellebrite failed to provide any information about the source device for the browsing history. As illustrated in Table 12 and 13 below, APOLLO explicitly states that the web history stemmed from the device. Cellebrite on the other hand, provided no such information, opening a potential vector for wrongful attributions.

[8] https://support.apple.com/en-kw/guide/icloud/mm5400ef10c4/1.0/icloud/1.0



29.06.2023 12:10:05	yr - Google-søk	Safari	/private/var/mobile/Library/Safari/History.db : 0x4F63 (Table: history_visits, history_items; Size: 122880 bytes)
29.06.2023 12:10:05	indisk restaurant skien - Google-søk	Safari	/private/var/mobile/Library/Safari/History.db : 0x4F22 (Table: history_visits, history_items; Size: 122880 bytes)
29.06.2023 12:10:06	indisk restaurant skien - Google-søk	Safari	/private/var/mobile/Library/Safari/History.db : 0x4EE1 (Table: history_visits, history_items; Size: 122880 bytes)
29.06.2023 12:10:08	indisk restaurant skien - Google-søk	Safari	/private/var/mobile/Library/Safari/History.db : 0x4EA1 (Table: history_visits, history_items; Size: 122880 bytes)
29.06.2023 12:10:10	indisk restaurant skien - Google-søk	KnowledgeC	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x25E283 (Table: ZOBJECT, ZSTRUCTUREDMETADATA; Size: 3833856 bytes)
29.06.2023 12:10:18	indisk restaurant skien - Google-søk	Safari	/private/var/mobile/Library/Safari/History.db : 0x4E61 (Table: history_visits, history_items; Size: 122880 bytes)
29.06.2023 12:10:28		KnowledgeC	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x25E102 (Table: ZOBJECT, ZSTRUCTUREDMETADATA; Size: 3833856 bytes)
29.06.2023 12:10:28	indisk restaurant porsgrunn - Google-søk	Safari	/private/var/mobile/Library/Safari/History.db : 0x4E1C (Table: history_visits, history_items; Size: 122880 bytes)
29.06.2023 12:10:28	indisk restaurant porsgrunn - Google-søk	Safari	/private/var/mobile/Library/Safari/History.db : 0x4DD7 (Table: history_visits, history_items; Size: 122880 bytes)
29.06.2023 12:10:35	indisk restaurant porsgrunn - Google-søk	KnowledgeC	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x266EA0 (Table: ZOBJECT, ZSTRUCTUREDMETADATA; Size: 3833856 bytes)
29.06.2023 12:10:37	indisk restaurant porsgrunn - Google-søk	Safari	/private/var/mobile/Library/Safari/History.db : 0x4D93 (Table: history_visits, history_items; Size: 122880 bytes)
29.06.2023 12:10:50	indisk restaurant porsgrunn - Google-søk	Safari	/private/var/mobile/Library/Safari/History.db : 0x4D4F (Table: history_visits, history_items; Size: 122880 bytes)
29.06.2023 12:10:51	indisk restaurant porsgrunn - Google-søk	Safari	/private/var/mobile/Library/Safari/History.db : 0x4D0A (Table: history_visits, history_items; Size: 122880 bytes)

Table 12: Detail level of web history in Cellebrite



29.06.2023 12:10:05	Safari Browsing	VISIT TIME: 29.06.2023 12:10:05 URL: https://www.google.no/search?q=indisk+restaurant+skien&client=safari&channel=iphone_bm&ei=alidZJuLD_qKxc8PkuePcA&oq=indisk+restaurant+skien&gs_lcp=ChNtb2JpbGUtZ3dzLXdpei1zZXJwEAMyBQgAEIAEMgYIABAWEB4yBggAEBYQHjIGCAAQFhAeMgUIABCiBDoKCAAQigUQsAMQQzoOCAAQgAQQsQMQgwEQsAM6FQguEIoFEMcBENEDEMgDELADEEMYAToRCC4QgAQQsQMQgwEQxwEQ0QM6CwgAEIoFELEDEIMBOgsILhCKBRCxAxCDAToOCC4QgAQQsQMQxwEQ0QM6BwgAEIoFEEM6EwguEIoFELEDEIMBEMcBENEDEEM6EAguEIoFELEDEMcBENEDEEM6DQgAEIoFELEDEIMBEEM6CwgAEIAEELEDEIMBOgsILhCABBCxAxCDAToLCC4QgAQQxwEQrwE6CAguEIAEELEDOggIABCABBCxAzoLCC4QrwEQxwEQgAQ6CwgAEIAEELEDEMkDOggIABCKBRCSAzoICAAQgAQQkgM6DggAEIAEELEDEIMBEMkDSgQIQRgBUPg1WOmBAWDhhQFoAnAAeACAAb0BiAHmEplBBDkuMTSYAQCgAQGwAQDAAQHIARHaAQQIARgI&sclient=mobile-gws-wiz-serp VISIT COUNT: 2 TITLE: indisk restaurant skien - Google-sÃ¸k ICLOUD SYNC: VISITED FROM THIS DEVICE LOAD SUCCESSFUL: 1 VISIT ID: 143 REDIRECT SOURCE: None REDIRECT DESTINATION: None HISTORY ITEM ID: 143	History.db
29.06.2023 12:10:05	Safari Browsing	VISIT TIME: 29.06.2023 12:10:05 URL: https://www.google.no/search?q=yr&client=safari&channel=iphone_bm&ei=xC4SYYLhDLDIrgTblaGYAw&oq=yr&gs_lcp=ChNtb2JpbGUtZ3dzLXdpei1zZXJwEAMyCggAELEDEIMBEEMyBAgAEEMyCwgAEIAEELEDEIMBMgoIABCxAxCDARBDMgoIABCxAxCDARBDMgoIABCxAxCDARBDMgsIABCABBCxAxCDATILCAAQgAQQsQMQgwE6BggAEAcQHjoFCAAQgAQ6CwgpEIAEELEDEIMBOglIKToTCCkQsQMQgwEQxwEQ0QMQQxCTAjoKCCkQsQMQgwEQQzoRCC4QgAQQsQMQgwEQxwEQ0QM6CwguEIAEELEDEIMBOggIABCABBCxA1CfFlirGGCXH2gAcAB4AIABZIgBnwKSAQMyLjGYAQCgAQGwAQnAAQE&sclient=mobile-gws-wiz-serp VISIT COUNT: 3 TITLE: yr - Google-sÃ¸k ICLOUD SYNC: VISITED FROM THIS DEVICE LOAD SUCCESSFUL: 1 VISIT ID: 142 REDIRECT SOURCE: None REDIRECT DESTINATION: None HISTORY ITEM ID: 142	History.db

Table 13: Detail level of web history in APOLLO

5.2.6 Task 6: Original of photo

> ”A photo with timestamp 18:07 has been found on Berit’s iPhone. The image is unclear, but a person’s legs appears to be in the background. The lead investigator would like you to investigate whether the photo was received or taken on the device. This could indicate whether she was with someone in the time before her death.”

In the sixth task, the respondents were asked to determine the origin of a blurry photo, with what seems to be a person in the background.

The specific photo was taken during the scenario to establish that the victim was, in fact, with someone in the time before her death. As all participants were confined to an Excel spreadsheet for their analysis, they had to rely on metadata to assess its origin.

Determining the origin of media files is a domain with several facets. However, in this task, it was expected that the participants established sequences of events that either supported or refuted that the photo was taken with the device. Activities of interest would be use of the camera application, activation of camera, in addition to information parsed from the devices photo database.

APOLLO:
All participants in the APOLLO group concluded that the photo was taken with the device, thereby establishing that the victim could have been with someone in the time before her death.

Despite that the tool provided detailed descriptive information about the photo, they based their conclusion on an attribute stating "Made/Saved with this device". The respondents' interpretation of this descriptor seems to be literal, as my own testing reveals that the same descriptor appears in several different scenarios: When a photo is taken with the device, when a screenshot is taken, and when a photo is saved to the photo album through a third-party application.

One respondent included additional information, demonstrating a correlation between the use of the device's camera application and the photo's timestamp. However, the activity selected by this participant only showcased the camera application's start time, missing out on the duration of the usage. In this case, use of APOLLO's activity "Application in Focus" would have added additional information, establishing that the camera application was active at the same time frame as the photo was saved to the device.

Two of the participants provided further arguments, stating that the photo most likely were taken with the device, as it was saved in Apple's HEIC format. However, none of the participants built a solid sequence of events to supporting their hypothesis.

Cellebrite group:
The group failed to produce any descriptive information about the photo. Instead, the they relied on application usage to establish that the camera application had been open in the same time frame as the photo was saved. Two of the respondents added additional data regarding activation of the devices camera.

One of the participants expressed uncertainty whether their selected activity really showed the device's camera state. Another participant highlighted that the camera state indicator did not give any contextual information, indicating if a photo had being taken or not. Despite being able to infer that a photo might have been taken, the tool failed to provide any information about the actual photo.

A major disadvantage for the group, was a lack of any data parsed from the iOS photo database. Therefore, the existence of the photo had to be inferred based on information given in the assignment, and metadata such as application usage and camera activation.

General findings:
The APOLLO group had a major advantage, as detailed metadata was parsed from the iOS photo database. Despite having this advantage, the entire group made false assumptions regarding the descriptor "Made/Saved with this device". The reason behind this misinterpretation seems to be a reliance on literal interpretations where in-depth knowledge is lacking.

The Cellebrite group failed to produce any information about the photo itself. Therefore, the participants had to substantiate its existence based on application usage and camera activation. It is however unlikely that they would have recognised its existence without the task description.

Note that all media files was removed from the data source prior to processing the data, as such timestamps from file listings were unavailable. However, both tools had access to the iOS photo database, containing detailed information regarding media files stored within the devices gallery.

Summary of errors:

- APOLLO: Three participants misinterpreted a descriptor within the photo activity due to naming conventions
- Cellebrite: Three participants failed to identify any image metadata due to lack of details

5.2.7 Task 7: Connection to charger

> ”A forensic investigator has noticed that the charging cable for Berit’s iPhone is lying on the floor, approximately one meter away from the charging unit (which remains plugged in). This could be a possible situational trace, and the lead investigator has asked you to determine the last time the iPhone was plugged in.”

The seventh task presented a potential situational trace to the participants: the charging cable of the victim’s iPhone was found about 1 meter from the charge plug, which was still plugged into the socket. They were asked to assess the time for the last disconnect, as this could give valuable information regarding the victim’s last signs of life.

In the introduction to the scenario, it was noted that the victim’s phone was found by the police, approximately 130 meters from the residence (4.2). This specific action was performed to simulate the events that led up to this moment, namely that the perpetrator hastily unplugged the phone from the charger before leaving the residence.

APOLLO:
The APOLLO group used three different sets of activities to demonstrate the last time the device was unplugged: The first respondent used ”Device Plugin Status” in conjunction with the ”Battery Level”, the second respondent solely relied on ”Battery Level”, while last participant relied on ”Accessory Connection”.

Between these sets of activities, the use of ”Device Plugin Status” in conjunction with ”Battery Level” was the most informative, providing both the time for unplugging the device and devices battery percentage.

The use of ”Battery Level” alone was also a valid choice. Data within the activity was found to be sourced from two different database, KnowledgeC.db and CurrentPowerlog.PLSQL, whereby only the latter contained a charging status. However, this activity is updated in irregular intervals, which only produces an approximate timestamp dependent on its last update.

A closer examination of the activity ”Accessory Connected” reveal that is contains information about connected peripherals such as Bluetooth speakers and CarPlay connection through the lightning cable. In this instance, the activity selected by the participant was not related to the charging status as it represented the loss of connection to the Bluetooth speaker.

Cellebrite:
The Cellebrite group demonstrated a uniform approach to the assessment. All respondents correctly selected an activity with the descriptors ”Unplugged” and ”Plugged”, with timestamps that aligned with the ground truth (section 4.1.3). Despite being able to determine the correct timestamp, one participant noted that no information about the device’s battery level could be found.

In addition to the correct activity, one participant also included the Wi-Fi activity ”Connect” and ”Disconnect”, mistaking it for the device’s charging status.

General findings:
One participant in each group selected erroneous activities due to a literal interpretation of the activity names. In the case of APOLLO group, one participant selected an activity that showcased loss of connection to the Bluetooth speaker. In the Cellebrite group, one participant included Wi-Fi connection intervals mistaking them for charging statuses. This instance signifies the time the Cellebrite group misinterpreted the same Wi-Fi activity during the assignments.

Summary of errors:

- APOLLO: One participant misinterpreted an activity due to naming conventions
- Cellebrite: One participant misinterpreted an activity due to naming conventions

5.2.8 Task 8: Last activities

> ”The exact time of Berit’s death is uncertain. You are tasked to identify any activities that could indicate her last signs of life. This includes her last interactions with the device, as well as health data of relevance.”

In what might be the most demanding task, the respondents were required to select significant events and discern between data based on several criteria: Data generated on the device as a result of physical actions, data generated due to user interactions with the device, and data generated by the device itself.

Details presented in the scenario contained several hints as to the relevance of the task. This includes information that the device was found approximately 130 meters from the residence, that the charger cable seemed ripped out from the charger, and that the victim still wore her Apple Watch when she was found. Relevant data could in this regard be heart rate and steps, charging statuses, loss of Wi-Fi and Bluetooth connections, and application usage.

APOLLO:
One of the participants in the APOLLO group was unable to complete the assignment due to time constraints. Despite this, the participant did manage to complete Task 9 and the post-scenario interview. As the participant only selected two items their assessment of this particular task is excluded from any further evaluation.

The two remaining participants both utilized heart rate samples from the victim’s Apple Watch, commenting that the heart rate was elevated around 19:50. They also noted a sudden drop in heart rate before an abundance of further samples. Both participants drew a hypothesis based on this behavior, and attributing it to a potential physical confrontation.

They also noted a rapid increment in steps in the aftermath of 19:50. One of them commented on the possibility that all recorded actions, in the aftermath of the sudden drop in heart rate, could possibly be attributed to the perpetrator. The same respondent also noted that all data generated after 22:00, most certainly was generated by the police’s handling of the device.

The respondents also included other activities such as activity level, camera usage and some location data, but failed to include them in any overall hypothesis.

Both participants produced a well-rounded assessment that aligned with several aspects of the ground truth. Despite accounting for several aspects of the ground truth, they failed to incorporate activities used in earlier assignments such as loss of Wi-Fi and Bluetooth connections.

Cellebrite:
The first respondent in the Cellebrite group highlighted that the victim’s heart rate was exaggerated at 20:00 reading 158BPM. They noted that no further specifics could be given, due to the tool’s aggregation of health data. As a response to this, they suggested a manual analysis of the health database to obtain further specifics. In addition, they noted usage of the camera application, but did not continue with any further inquiry. The respondent was not able to provide any definite hypothesis, suggesting that a manual investigation was required. As the victim was found at 22:00, they inferred that any usage after this point must be attributed to the police handling the device.

The second participant was uncertain how to interpret the to/from timestamps for the health data. They noted that data recorded between 18:00 and 20:00 were significantly greater than earlier in the day, and that max heart rate was 158. They found no new recordings after 21:00, but were uncertain if the heart rate of 158 happened between 19:00-20:00 or 20:00-21:00. Additionally, they noted the use of Tinder at 19:44, suggesting that this event could have been the root of a physical confrontation.

The last respondent recognized that the heart rate was elevated sometimes around 20:00 and that no further health samples were recorded in the aftermath. As the next health data recording at 21:00 did not contain any steps or heart rate samples, the respondent suggested that the victim died sometime between 20:00 and 21:00. They inferred that last trace of active usage appeared at

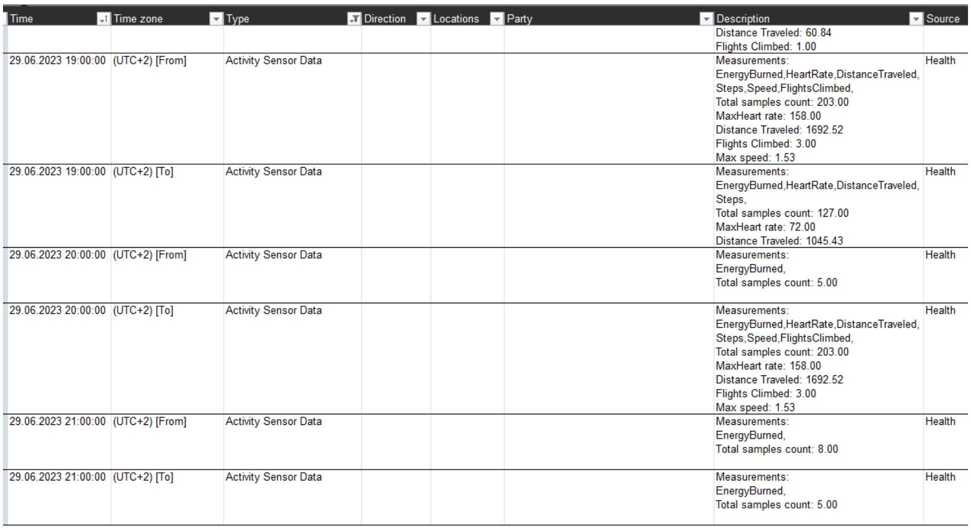
20:00:08 with an unlock of the device.

The Cellebrite group experienced significant difficulties due to the tool's aggregation of health-related data. As an example: While all of the participants managed to identify that the heart rate had been elevated, they were only able to pinpoint that it had been elevated sometimes within the span an hour.

Health data within Cellebrite's timeline is aggregated into a single record on an hourly basis. This single record consist of an aggregated summary that includes the total number of steps and flights climbed, movement in meters, in addition to a single value stating the max heart rate within the hour. This practice of data aggregation denied the participants' access to detailed records as well as specific timestamps.

In addition to the above-mentioned difficulties, they experienced confusion when interpreting the start and end times of each aggregated time frame: One participant misinterpreted the timestamp, stating that the elevated heart rate occurred exactly at 20:00, due to confusing the end time for the actual time this heart rate was registered. The two remaining participants understood the concept of aggregation, but experienced uncertainty when interpreting when the actual periods had started and ended. As of this, they were unable to determine whether the aggregated data represented values between 19:00 - 20:00 or 20:00 - 21:00.

The respondents uncertainties regarding these timestamps appears to be directly related to the tool's presentation layer. As demonstrated in Table 14, the data are presented with two rows each hour, whereby the rows with identical timestamps are seemingly randomly arranged. In addition, the participants found the duplicate values for each respective time period to be confusing.



Time	Time zone	Type	Direction	Locations	Party	Description	Source
						Distance Traveled: 60.84 Flights Climbed: 1.00	
29.06.2023 19:00:00	(UTC+2) [From]	Activity Sensor Data				Measurements: EnergyBurned,HeartRate,DistanceTraveled, Steps,Speed,FlightsClimbed, Total samples count: 203.00 MaxHeart rate: 158.00 Distance Traveled: 1692.52 Flights Climbed: 3.00 Max speed: 1.53	Health
29.06.2023 19:00:00	(UTC+2) [To]	Activity Sensor Data				Measurements: EnergyBurned,HeartRate,DistanceTraveled, Steps, Total samples count: 127.00 MaxHeart rate: 72.00 Distance Traveled: 1045.43	Health
29.06.2023 20:00:00	(UTC+2) [From]	Activity Sensor Data				Measurements: EnergyBurned, Total samples count: 5.00	Health
29.06.2023 20:00:00	(UTC+2) [To]	Activity Sensor Data				Measurements: EnergyBurned,HeartRate,DistanceTraveled, Steps,Speed,FlightsClimbed, Total samples count: 203.00 MaxHeart rate: 158.00 Distance Traveled: 1692.52 Flights Climbed: 3.00 Max speed: 1.53	Health
29.06.2023 21:00:00	(UTC+2) [From]	Activity Sensor Data				Measurements: EnergyBurned, Total samples count: 8.00	Health
29.06.2023 21:00:00	(UTC+2) [To]	Activity Sensor Data				Measurements: EnergyBurned, Total samples count: 5.00	Health

Table 14: Activity sensor data in Cellebrite

Due to lack of specifics, the group refrained from any conclusions regarding the elevated heart rate. Only one participant contributed with information of significance for the ground truth, in addition to the aggregated health data. This participant noted that the Tinder application had been open, and theorized that this could have escalated a conflict between the victim and the perpetrator.

General findings:
Only two participants in the APOLLO group were able to complete the assignment. These two participants produced well rounded assessments that accounted for several aspects of the ground truth. Due to the tool's granular presentation of health data, they were able to hypothesize that the sudden elevation in heart rate, followed by an absence of further readings, might be the indication of a physical conflict. One of the participants also suggested that subsequent steps possibly could be attributed to the perpetrator taking possession of the device.

In contrast, the Cellebrite group experienced significant difficulties due to the tool's aggregation of health data. The group not only lacked access to necessary details, but experienced confusion when interpreting the start and end times for each aggregated period.

Pinpointing these events would undoubtedly be of grave importance in a real-life investigation. As shown, differences in outcome can to a large degree be attributed to the tools presentation layers. While APOLLO presented a straightforward tables containing all health samples, Cellebrite aggregated health data on an hourly basis, providing a single summarizing value for each distinct health type. Whereas the APOLLO group was able to construct narratives based granular data, the Cellebrite group refrained from any theorizing due to a lack of data.

Summary of errors:

- Cellebrite: Three participants failed find necessary specifics regarding health data due to data aggregation
- Cellebrite: Three participants were confused due to the presentation of timestamps for aggregated periods

5.2.9 Task 9: Airplane mode

> ”When Berit’s mobile phone arrived at the police station, it was already set to airplane mode. It is somewhat unclear whether Berit herself, the police, or someone else activated the setting. The lead investigator wants you to determine the last time the device was set to airplane mode.”

In the last assignment, the participants were asked to determine the last time the device was set to airplane mode. The task was created to simulate an event where lack of documentation led to confusion whether it was the police who set the device into airplane mode upon seizure or not.

APOLLO:
All respondents in the APOLLO group were able to pinpoint the time the device was set to airplane mode.

Cellebrite:
As with APOLLO, all respondents in the Cellebrite group were able to pinpoint the time the device was set to airplane mode.

General findings:
As expected, none of the groups experienced challenges in this assignment. All participants solved the assessment in a similar way, highlighting one single activity related to the status of the device’s airplane mode.

5.3 Post-scenario interviews

As apposed to the first part of the interview which focused on the participants background, this section explores various aspects of their experiences while working on the assignments.

5.3.1 Technical challenges

To control for the integrity of the experiment participants were asked if they had encountered any technical difficulties:

Only one participant reported a technical difficulty. This issue was related to an inherent limitation in Microsoft Excel that only allows for filtering on dates within the first 50 thousand rows of a dataset. Despite reporting this limitation, the participant was not affected as they circumvented the limitation by filtering on additional rows.

5.3.2 Missing data

The respondents were asked if they encountered difficulties when searching for any specific information, and if any expected data were missing from the dataset:

APOLLO:
Two of the participants in the APOLLO group expressed confusion over unfamiliar BSSIDs while assessing Wi-Fi connections in Task 2 (5.2.2). More specifically, they found what looked like BSSIDs, but were unable to convert as they were stored in an unfamiliar format.

The last participant expressed that although the APOLLO did not lack any specific data, they would rather review the data in a graphical user interface similar to that of Cellebrite Physical Analyzer.

Cellebrite:
Two of the respondents expressed frustration as the entire dataset were lacking in details. One expressed it as "only seeing the tip of the iceberg", while the other stated that "it was like only reading the headlines of the news". Due to the general lack of details, one participant was concerned that they might have fallen into biases during the assignments.

The third participant specifically highlighted health data, Wi-Fi, and Bluetooth as areas where they expected more details. They mentioned how the tools' aggregation of data sometimes made it difficult to pinpoint any specifics. Regarding Wi-Fi and Bluetooth connections, they had expected that both the names and identifiers of each connection would have appeared.

5.3.3 Interpretation

The respondents were asked if they encountered any data that was difficult to interpret, and were asked to elaborate on any specific challenges they encountered:

APOLLO:
The first participant had the initial experience of an overwhelming amount of data. Nevertheless, they quickly developed a technique for filtering that reduced it to a manageable level. They were satisfied with the detail level as it provided most of the details needed throughout the assignments. In some instances, similar activity names made it difficult to differentiate between the meaning of activities, especially when they showcased similar data from different databases.

The second participant experienced similar challenges when differentiating between activities with similar names. However, they attributed this confusion to inexperience rather than the tool itself.

The last participant experienced Task 8 (5.2.8) as especially challenging. They found that some of the data were hard to contextualize, especially when trying to distinguish between system and user generated data. Despite these challenges, they found that the tool provided a detail level that far surpassed that of Cellebrite.

Cellebrite:
The first respondent emphasized that the tool showcased Wi-Fi connections with identical timestamps and that they did not show the time of disconnection. They mentioned that iOS regularly gathers Wi-Fi information from Apple's servers, and warned that this is a common source of misinterpretation.

They pointed out that activities such as Wi-Fi connections, charging statuses, and camera usage all lacked important details that are present in APOLLO.

They stated that the tools naming conventions were a major source of misinterpretation. Terms such as "on/off", "connect/disconnect", often appeared without proper context, which led to guesswork regarding their meaning.

The second participant found it hard to distinguish between activities that contained terms such as "on/off", "plugged/unplugged", "connect/disconnect", "speaker", as they often lacked descriptive information. To interpret these instances, they had to refer to the source file column in order to guess their meaning.

They also experienced confusion when trying to determine when an activity had started and ended. The root cause of this problem was that timestamps were segregated into two separate rows with many rows in between. As a result, they had to search the table, trying to match corresponding start and end timestamps, an exercise that required extra effort.

They were also confused as to when aggregated time periods had started and ended. This was especially true when interpreting health data, as steps, distance, floors, and heart rate, were summarized into two rows with identical data. Since the previous end period had an identical timestamp to the new start time, and these rows appeared in random order, it was difficult to understand which hour each row was referring to.

The last participant faced difficulties interpreting Wi-Fi connections due to lack of necessary details. In relation to the information needs, they felt that the available data only was sufficient as a starting point for further analysis, and would have needed access to the source material for a proper analysis. They had an overall feeling that the tool lacked the necessary details to perform this type of analysis.

5.3.4 Accuracy and reliability

The respondents were asked if they considered the accuracy and reliability of the data, and if they had noted anything they felt the need to verify or confirm with other methods:

APOLLO:
The first respondent felt the need to verify timestamps in some of the health-related data they had tagged, not because they had distrust in the accuracy of the data, but to ensure that they had interpreted it correctly.

The second participants repeated the challenges they encountered when attempting to convert Wi-Fi BSSIDs in Task 2 (5.2.2), and felt to need to for more research on the subject due to its unfamiliar format.

The last participant felt the need to perform further analysis in Task 8 (5.2.8), as they experienced uncertainties when distinguish between system and user generated data. They had felt more confidence is they could have cross-validated their results with Cellebrite.

Cellebrite:
The first respondent felt that it was a straightforward task to interpret data such as browsing history and SMS messages. Activities such as charging statuses, activation of flight mode and camera usage were more problematic, as their ambiguous naming conventions required prior knowledge to interpret. Despite that they had prior knowledge of several of the activities, they felt the need to cross-verify their results with another tool.

The second participant mentioned challenges that arose when trying to interpret the meaning of ambiguous descriptors, and when interpreting segregated time frames. As such they felt a need to explore the data further to gain more knowledge about these activities.

The third participant found that the dataset they were provided had insufficient details to draw any clear conclusions. In a real-life setting, they would have been hesitant to write a forensic report on activity data based on these results. They mentioned that mitigating the lack of details would require access to the source data, in addition to a secondary tool.

5.3.5 Tool selection

The participants were asked if they had any thoughts about using different tools, and whether this would have influenced their findings:

APOLLO:
The first participant believed that their assignments would have been more challenging with a dataset from Cellebrite. This statement was based on their prior knowledge of the tool and how it lacks details related to activity data. Although they suspected that this characteristic could have led to some incorrect conclusions, this could be mitigated by verifying their results with two independent tools

The second participant explained that they found much more data in APOLLO compared to what they recall seeing in Cellebrite. They highlighted that Task 4 (5.2.4), 7 (5.2.7) and 8 (5.2.8) would have been more challenging with Cellebrite, especially when interpreting activities such as CarPlay and charging statuses.

The last participant did not have any specific thoughts whether the use of another tool would have influenced their findings.

Cellebrite:
The first participant felt that tools like APOLLO and iLEAPP would have provided them more insight into the data. They highlighted how these tools are generally more suitable for activity data, as data aggregation is a major disadvantage in Cellebrite. They specifically pointed out that they would have been able to find details of Wi-Fi connections in Task 2 (5.2.2) had they used either iLEAPP or APOLLO.

The second participant stated that they favored other tools when reviewing activity data, and that an alternative tool would have led to fewer errors.

The last participant suggested that they might have been able to fully analyze the information if they had access to the full version of Cellebrite Physical Analyzer, and were uncertain if an export from another tool would have yielded better results. They compared the results they were given to only having ”a single piece of a puzzle”.

5.3.6 Competency

The respondents were asked if any special skills or prior knowledge was needed to analyze the information needs:

APOLLO
All of the respondents felt that it was necessary with prior knowledge of distinct artifacts to answer the information needs. They emphasized that an the importance of in-depth understanding to differentiate between activities.

One participant added that even though they had limited knowledge of the tool, they felt that their prior knowledge was sufficient to infer meaning to most activities.

Another participant highlighted competency in filtering as a critical skill to gain.

The last participant emphasized knowledge of testing and post-test verification as a critical skill, had it been a real-life investigation.

Cellebrite:
The first respondent highlighted knowledge of technical analysis as an essential factor for understanding the data. They emphasized that prior experience with open-source analysis tools had helped them gain additional competency, as the tools openness allowed them to study how the data was parsed. The use of open-source tools had in turn helped them to better understand the data in Cellebrite, especially because the tool reduces complexity.

The second participant believed that their assessments would have produced fewer fault with more prior knowledge. Despite having considerable experience with log file analysis, they found data presented by Cellebrite more difficult to interpret compared to past experiences. They stated that even with more prior knowledge, they would still have struggled with several assignments due to a fundamental lack of details.

The last participant highlighted general knowledge of computer systems as a necessity. They also mentioned the need for competency in filtering data, as these forms of analysis consists of large amount of data. Despite that they were proficient in filtering data, they still struggled to make sense of all the activity types.

5.3.7 Additional comments

The participants were asked if they had any questions or something to add to the cases and interviews:

All of the respondents found their participation in the experiment interesting. There were no questions related to the actual implementation, but all participants were curious to different aspects of the ground truth.

6 Discussion

6.1 Factors of the tools interpretation layers

The following section contains a synthesis of insights gathered through the preceding analysis and discussion. It focuses on elements within the tools presentation layers that were found to affect different aspects of the assignments. The discussion leads back to the research question and aims to answer if factors such as data aggregation, presentation of timestamps and naming conventions affect the conclusions of investigators.

6.1.1 Naming conventions

Efficient use of analysis tools is dependent on clear and concise naming conventions. This is especially true with tools that present data from a multitude of sources and when utilized by less experienced users. Naming conventions should be intelligible across all categories, activities, and descriptors to ensure that users understand the meaning and context behind the data.

Several sources of errors related to naming conventions were discussed in the preceding sections. To synthesize the findings further, a distinction has been made between descriptors, activities, and categories:

Descriptors:
Descriptors refer to the labeling of specific attributes within the presentation layer. Examples of this could be field or column names, or in the context of communications, identifiers that describe individual attributes, such as "name" and "phone number". As the name implies, the purpose of descriptors is to describe the meaning and context of the data.

Activities:
Activities, on the other hand, refer to groups of data parsed from specific sources. Examples include communications from specific applications, or a single type of entity within the device's health data. The purpose of segmenting data into activity types is to allow the user to easily distinguish between different types of data.

Categories:
Categories serve as unified collections of activities, segregated by themes such as communications or health data. As an example, one category could consist of multiple related activities, such as messages from several communication applications, or collections of health related activities. The purpose of placing data within specific categories is to offer a consolidated view of similar data.

Findings:
The participants displayed a tendency to interpret activity names literally, a tendency that led to

multiple misinterpretations throughout the assignments. Although misinterpretations were found in both groups, they occurred much more frequently in Cellebrite compared to APOLLO.

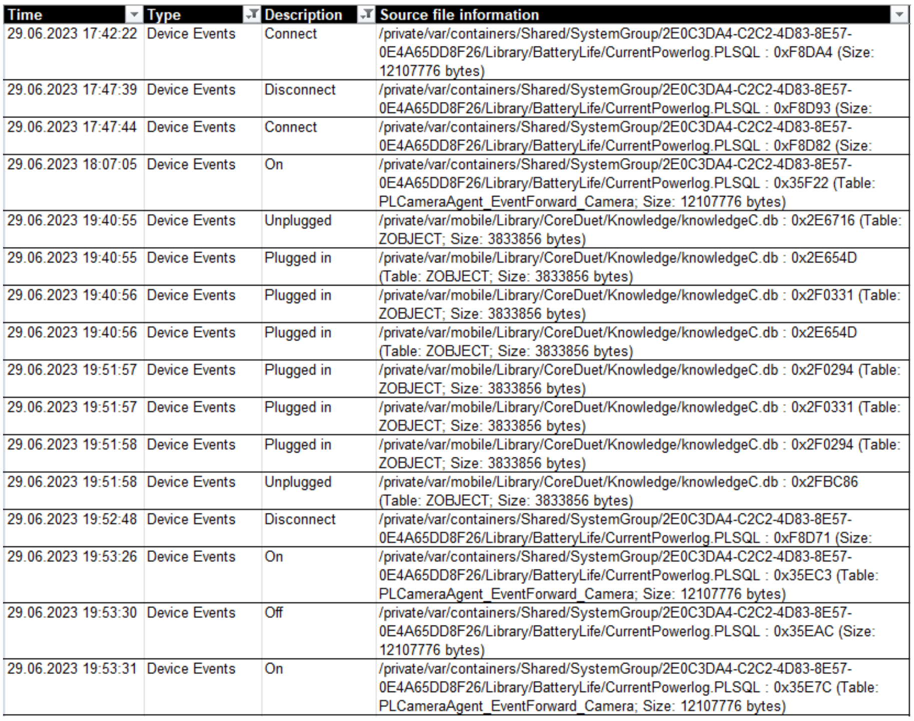
The Cellebrite group experienced repeated challenges due to the tool's naming conventions, especially when interpreting activities within the category "Device events". The category contains a mixture of activities, such as Wi-Fi and Bluetooth connections, as well as statuses for charging, lock screen, and CarPlay usage. Activities found within the category were found to be labeled with descriptors such as "Connect/Disconnect", "On/Off", "Plugged in/Unplugged", offering little context other than a source file reference.

One of the activities, "Connect/Disconnect", was misinterpreted several times throughout the assessments. In one instance, it was interpreted as the device's charging status, and in another instance as connections to Bluetooth devices. Although the activity actually consists of inferred Wi-Fi connection intervals, none of the participants evaluated it during the assessment of Wi-Fi connections in Task 2 (5.2.2). Adding to this confusion, they interpreted the category "Wireless Networks" literally, and misinterpreted it as Wi-Fi connections, while it actually showcased harvested Wi-Fi locations without coordinates (5.2.2).

The participants echoed these findings in the post-scenario interviews, expressing that the labels were experienced as vague and ambiguous. This led to recurring confusions, a situation that most likely could have been avoided had the tool provided sufficient details. Descriptors such as "on/off", "connect/disconnect", and "plugged/unplugged" were highlighted due to their ambiguity, which made it difficult to distinguish one activity from another (5.3.3).

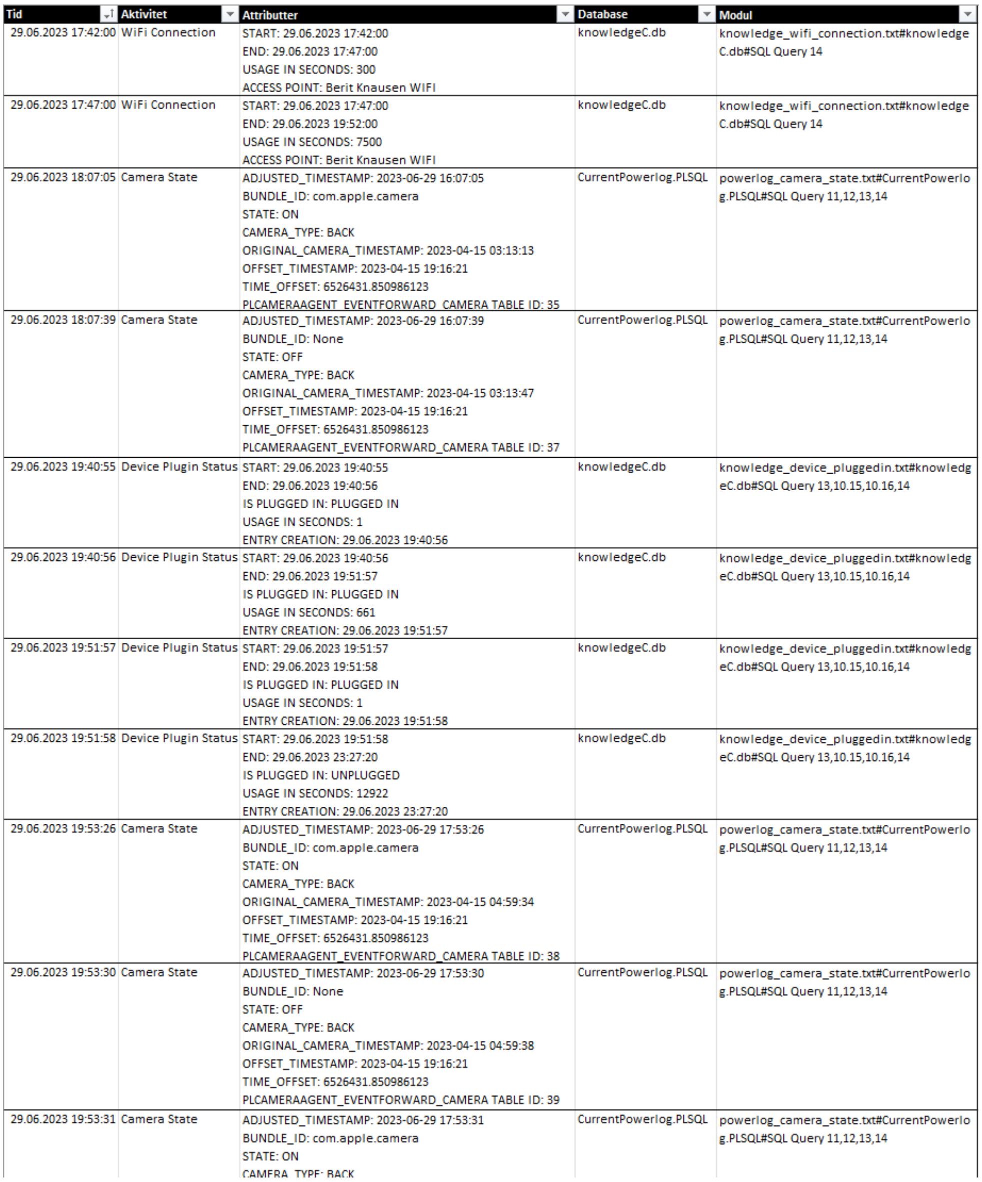
The APOLLO group experienced significantly fewer instances of confusions due to naming conventions. The tool implements a different approach than Cellebrite, in that it does not provide any overarching categories with similar activities. The tool depends less on data aggregation, resulting in more granularity with respect to the number of activities compared to Cellebrite. Some of the participants experienced the tool's more than 200 activities as overwhelming, struggling to differentiate between similar sounding activity names. Despite experiencing some confusion, only one significant instance was found, leading a participant to miss outgoing SMS communications (5.2.1).

The difference in naming conventions between the tools is illustrated in Table 15 and 16 below.



Time	Type	Description	Source file information
29.06.2023 17:42:22	Device Events	Connect	/private/var/containers/Shared/SystemGroup/2E0C3DA4-C2C2-4D83-8E57-0E4A65DD8F26/Library/BatteryLife/CurrentPowerlog.PLSQL : 0xF8DA4 (Size: 12107776 bytes)
29.06.2023 17:47:39	Device Events	Disconnect	/private/var/containers/Shared/SystemGroup/2E0C3DA4-C2C2-4D83-8E57-0E4A65DD8F26/Library/BatteryLife/CurrentPowerlog.PLSQL : 0xF8D93 (Size:
29.06.2023 17:47:44	Device Events	Connect	/private/var/containers/Shared/SystemGroup/2E0C3DA4-C2C2-4D83-8E57-0E4A65DD8F26/Library/BatteryLife/CurrentPowerlog.PLSQL : 0xF8D82 (Size:
29.06.2023 18:07:05	Device Events	On	/private/var/containers/Shared/SystemGroup/2E0C3DA4-C2C2-4D83-8E57-0E4A65DD8F26/Library/BatteryLife/CurrentPowerlog.PLSQL : 0x35F22 (Table: PLCameraAgent_EventForward_Camera; Size: 12107776 bytes)
29.06.2023 19:40:55	Device Events	Unplugged	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x2E6716 (Table: ZOBJECT; Size: 3833856 bytes)
29.06.2023 19:40:55	Device Events	Plugged in	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x2E654D (Table: ZOBJECT; Size: 3833856 bytes)
29.06.2023 19:40:56	Device Events	Plugged in	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x2F0331 (Table: ZOBJECT; Size: 3833856 bytes)
29.06.2023 19:40:56	Device Events	Plugged in	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x2E654D (Table: ZOBJECT; Size: 3833856 bytes)
29.06.2023 19:51:57	Device Events	Plugged in	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x2F0294 (Table: ZOBJECT; Size: 3833856 bytes)
29.06.2023 19:51:57	Device Events	Plugged in	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x2F0331 (Table: ZOBJECT; Size: 3833856 bytes)
29.06.2023 19:51:58	Device Events	Plugged in	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x2F0294 (Table: ZOBJECT; Size: 3833856 bytes)
29.06.2023 19:51:58	Device Events	Unplugged	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x2FBC86 (Table: ZOBJECT; Size: 3833856 bytes)
29.06.2023 19:52:48	Device Events	Disconnect	/private/var/containers/Shared/SystemGroup/2E0C3DA4-C2C2-4D83-8E57-0E4A65DD8F26/Library/BatteryLife/CurrentPowerlog.PLSQL : 0xF8D71 (Size:
29.06.2023 19:53:26	Device Events	On	/private/var/containers/Shared/SystemGroup/2E0C3DA4-C2C2-4D83-8E57-0E4A65DD8F26/Library/BatteryLife/CurrentPowerlog.PLSQL : 0x35EC3 (Table: PLCameraAgent_EventForward_Camera; Size: 12107776 bytes)
29.06.2023 19:53:30	Device Events	Off	/private/var/containers/Shared/SystemGroup/2E0C3DA4-C2C2-4D83-8E57-0E4A65DD8F26/Library/BatteryLife/CurrentPowerlog.PLSQL : 0x35EAC (Size: 12107776 bytes)
29.06.2023 19:53:31	Device Events	On	/private/var/containers/Shared/SystemGroup/2E0C3DA4-C2C2-4D83-8E57-0E4A65DD8F26/Library/BatteryLife/CurrentPowerlog.PLSQL : 0x35E7C (Table: PLCameraAgent_EventForward_Camera; Size: 12107776 bytes)

Table 15: Device events in Cellebrite



Tid	Aktivitet	Attributter	Database	Modul
29.06.2023 17:42:00	WiFi Connection	START: 29.06.2023 17:42:00 END: 29.06.2023 17:47:00 USAGE IN SECONDS: 300 ACCESS POINT: Berit Knausen WIFI	knowledgeC.db	knowledge_wifi_connection.txt#knowledgeC.db#SQL Query 14
29.06.2023 17:47:00	WiFi Connection	START: 29.06.2023 17:47:00 END: 29.06.2023 19:52:00 USAGE IN SECONDS: 7500 ACCESS POINT: Berit Knausen WIFI	knowledgeC.db	knowledge_wifi_connection.txt#knowledgeC.db#SQL Query 14
29.06.2023 18:07:05	Camera State	ADJUSTED_TIMESTAMP: 2023-06-29 16:07:05 BUNDLE_ID: com.apple.camera STATE: ON CAMERA_TYPE: BACK ORIGINAL_CAMERA_TIMESTAMP: 2023-04-15 03:13:13 OFFSET_TIMESTAMP: 2023-04-15 19:16:21 TIME_OFFSET: 6526431.850986123 PLCAMERAAGENT_EVENTFORWARD_CAMERA TABLE ID: 35	CurrentPowerlog.PLSQL	powerlog_camera_state.txt#CurrentPowerlog.PLSQL#SQL Query 11,12,13,14
29.06.2023 18:07:39	Camera State	ADJUSTED_TIMESTAMP: 2023-06-29 16:07:39 BUNDLE_ID: None STATE: OFF CAMERA_TYPE: BACK ORIGINAL_CAMERA_TIMESTAMP: 2023-04-15 03:13:47 OFFSET_TIMESTAMP: 2023-04-15 19:16:21 TIME_OFFSET: 6526431.850986123 PLCAMERAAGENT_EVENTFORWARD_CAMERA TABLE ID: 37	CurrentPowerlog.PLSQL	powerlog_camera_state.txt#CurrentPowerlog.PLSQL#SQL Query 11,12,13,14
29.06.2023 19:40:55	Device Plugin Status	START: 29.06.2023 19:40:55 END: 29.06.2023 19:40:56 IS PLUGGED IN: PLUGGED IN USAGE IN SECONDS: 1 ENTRY CREATION: 29.06.2023 19:40:56	knowledgeC.db	knowledge_device_pluggedin.txt#knowledgeC.db#SQL Query 13,10.15,10.16,14
29.06.2023 19:40:56	Device Plugin Status	START: 29.06.2023 19:40:56 END: 29.06.2023 19:51:57 IS PLUGGED IN: PLUGGED IN USAGE IN SECONDS: 661 ENTRY CREATION: 29.06.2023 19:51:57	knowledgeC.db	knowledge_device_pluggedin.txt#knowledgeC.db#SQL Query 13,10.15,10.16,14
29.06.2023 19:51:57	Device Plugin Status	START: 29.06.2023 19:51:57 END: 29.06.2023 19:51:58 IS PLUGGED IN: PLUGGED IN USAGE IN SECONDS: 1 ENTRY CREATION: 29.06.2023 19:51:58	knowledgeC.db	knowledge_device_pluggedin.txt#knowledgeC.db#SQL Query 13,10.15,10.16,14
29.06.2023 19:51:58	Device Plugin Status	START: 29.06.2023 19:51:58 END: 29.06.2023 23:27:20 IS PLUGGED IN: UNPLUGGED USAGE IN SECONDS: 12922 ENTRY CREATION: 29.06.2023 23:27:20	knowledgeC.db	knowledge_device_pluggedin.txt#knowledgeC.db#SQL Query 13,10.15,10.16,14
29.06.2023 19:53:26	Camera State	ADJUSTED_TIMESTAMP: 2023-06-29 17:53:26 BUNDLE_ID: com.apple.camera STATE: ON CAMERA_TYPE: BACK ORIGINAL_CAMERA_TIMESTAMP: 2023-04-15 04:59:34 OFFSET_TIMESTAMP: 2023-04-15 19:16:21 TIME_OFFSET: 6526431.850986123 PLCAMERAAGENT_EVENTFORWARD_CAMERA TABLE ID: 38	CurrentPowerlog.PLSQL	powerlog_camera_state.txt#CurrentPowerlog.PLSQL#SQL Query 11,12,13,14
29.06.2023 19:53:30	Camera State	ADJUSTED_TIMESTAMP: 2023-06-29 17:53:30 BUNDLE_ID: None STATE: OFF CAMERA_TYPE: BACK ORIGINAL_CAMERA_TIMESTAMP: 2023-04-15 04:59:38 OFFSET_TIMESTAMP: 2023-04-15 19:16:21 TIME_OFFSET: 6526431.850986123 PLCAMERAAGENT_EVENTFORWARD_CAMERA TABLE ID: 39	CurrentPowerlog.PLSQL	powerlog_camera_state.txt#CurrentPowerlog.PLSQL#SQL Query 11,12,13,14
29.06.2023 19:53:31	Camera State	ADJUSTED_TIMESTAMP: 2023-06-29 17:53:31 BUNDLE_ID: com.apple.camera STATE: ON CAMERA_TYPE: BACK	CurrentPowerlog.PLSQL	powerlog_camera_state.txt#CurrentPowerlog.PLSQL#SQL Query 11,12,13,14

Table 16: Device Events in APOLLO

Summary:
As revealed, naming conventions played an essential role in the users data comprehension. The errors that were found seem partly based on the participants tendency to interpret labels literally, and partly based on a lack of intelligible details. Confusions related to naming conventions appeared much more frequently in the Cellebrite group compared to APOLLO.

6.1.2 Details and contextual information

Assessing information needs at the activity level often requires insight into minuscule details supported by broad contextual information. Small elements and specifics can be essential in determining whether a particular information need can be answered or not. Lack of specifics, such as identifiers or references to source devices, could not only skew an investigation, but lead to erroneous attributions regarding the origin of the data. In addition, access to details and rich contextual information could support an investigator by broadening their understanding of particular artifacts, in addition to highlighting additional use cases.

The following section continues with the synthesis of the preceding discussions, now focusing on instances where lack of details and contextual information were found to affect assessments. Two main areas were identified:

- Instances where the user successfully identified an activity, but lack of essential details made them unable to answer specifics.
- Instances where lack of contextual information led participants to either overlook critical information or misinterpret their meaning.

In terms of data presentation, Cellebrite and APOLLO employ different strategies regarding the detail level presented to the end user. On one hand, Cellebrite was found to present parsed data in a simplistic manner, maintaining the number of attributes to a confined number. The reasoning behind this is somewhat unclear, but inherent limitations in the tool, as well as user friendliness could be factors of relevance. APOLLO, on the other hand, provided a more raw representation of its data, presenting attributes inherited directly from fields selected within its SQLite queries.

While both approaches arguably have their strengths and weaknesses, APOLLO's more raw presentation of data was favored during the assessments. The Cellebrite group experienced a lack of necessary specifics in several assignments:

Regarding instances where the user successfully identified an activity, but lack of essential details made them unable to answer specifics: In Task 3 (5.2.3), the group was unable to differentiate between connections to different Bluetooth devices due to a lack of specifics. This lack of specifics recursively affected Task 4 (5.2.4), as connections to the victim's Bluetooth speaker and the vehicle's Apple CarPlay were presented without unique identifiers such as device names and MAC addresses. The fact that the tool failed to present these specifics was surprising, as Cellebrite parses information about Bluetooth connections from KnowledgeC.db, a database that contains all necessary details. The APOLLO group, on the other hand, experienced no challenges answering the information needs, as identifiers such as device names, MAC addresses and connection intervals were readily available.

Regarding instances where lack of contextual information led participants to either overlook critical information or misinterpret their meaning: The Cellebrite group encountered somewhat similar circumstances during the assessment of Wi-Fi connections in Task 2 (5.2.2). However, this case differed in that the group was unable to identify the correct activity. While one participant refrained from answering the assignment, the remaining members wrongfully selected the activity "WIFI networks", which contained harvested Wi-Fi locations that are unrelated to the phone's Wi-Fi connections (Whiffin, 2022). The APOLLO group, on the other hand, had no issues determining the name and connection intervals for the network in question.

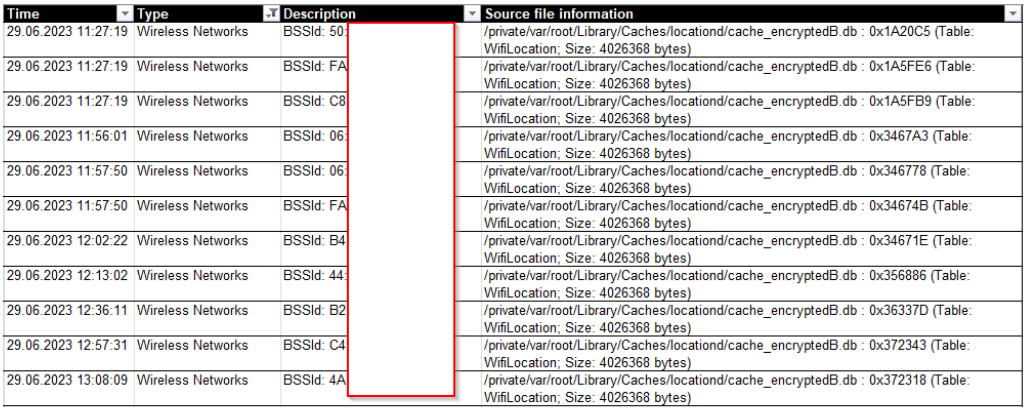
In the examples above, two variations of lacking details were demonstrated: In the first case the Cellebrite group was able to identify the correct activity, but failed to identify any specifics. In the second case, the group was unable to identify the correct activity and resorted to an erroneous activity instead. Even though only two examples were given here, the Cellebrite group experienced similar challenges across a series a assignments, demonstrating an inherent lack of necessary details rather than unique instances.

The findings aligned with the post-scenario interviews, where all participants in the Cellebrite Cellebrite group expressed concerns related to the tool's lack of details and contextual information. One participant described the detail level as "Only seeing the top of an iceberg", while another described it as "Reading news with only the headlines" (5.3). Common for the group was a feeling of having superficial access to the data, missing out on details that could have conveyed a deeper understanding of the events. The participants expressed that insufficient access to specifics details could lead to misinterpretations of evidence and incorrect conclusions.

Expanding on the previous section, challenges of ambiguous naming conventions seem closely intertwined with a lack of contextual information. Although labels such as "Connect/Disconnect", "On/Off", "Plugged in/unplugged" lacks in descriptive clarity, sufficient contextual information would most likely allowed for a sufficient inference of meaning.

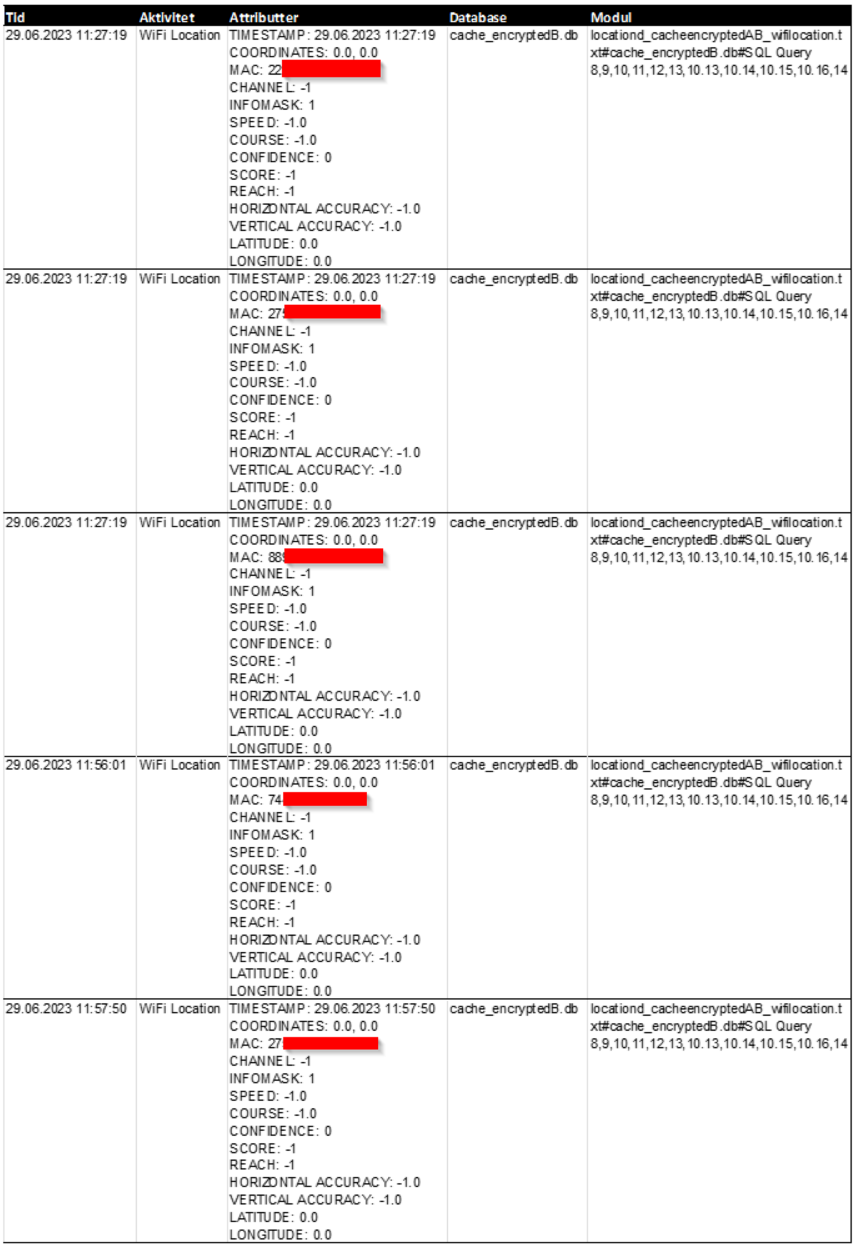
Table 17 contain an excerpt from the erroneous activity "Wireless Networks", selected by the

Cellebrite group in Task 2. While looking at the detail level provided by each of the tools, it becomes clear that the misinterpretations rests on a combination of ambiguous naming conventions and lack of contextual richness. Although not clearly stated by the tool, Cellebrite has divided harvested Wi-Fi locations into two categories, ”Wireless Network: Location” which contain harvested Wi-Fi locations with valid coordinates, and ”Wireless Networks” which contain harvested Wi-Fi locations without valid coordinates. This is more clearly illustrated in Table 18, showcasing how it was necessary to filter out all valid coordinates to obtain the same results as Cellebrite's activity ”Wireless Networks”.



Time	Type	Description	Source file information
29.06.2023 11:27:19	Wireless Networks	BSSId: 50[redacted]	/private/var/root/Library/Caches/locationd/cache_encryptedB.db : 0x1A20C5 (Table: WifiLocation; Size: 4026368 bytes)
29.06.2023 11:27:19	Wireless Networks	BSSId: FA[redacted]	/private/var/root/Library/Caches/locationd/cache_encryptedB.db : 0x1A5FE6 (Table: WifiLocation; Size: 4026368 bytes)
29.06.2023 11:27:19	Wireless Networks	BSSId: C8[redacted]	/private/var/root/Library/Caches/locationd/cache_encryptedB.db : 0x1A5FB9 (Table: WifiLocation; Size: 4026368 bytes)
29.06.2023 11:56:01	Wireless Networks	BSSId: 06[redacted]	/private/var/root/Library/Caches/locationd/cache_encryptedB.db : 0x3467A3 (Table: WifiLocation; Size: 4026368 bytes)
29.06.2023 11:57:50	Wireless Networks	BSSId: 06[redacted]	/private/var/root/Library/Caches/locationd/cache_encryptedB.db : 0x346778 (Table: WifiLocation; Size: 4026368 bytes)
29.06.2023 11:57:50	Wireless Networks	BSSId: FA[redacted]	/private/var/root/Library/Caches/locationd/cache_encryptedB.db : 0x34674B (Table: WifiLocation; Size: 4026368 bytes)
29.06.2023 12:02:22	Wireless Networks	BSSId: B4[redacted]	/private/var/root/Library/Caches/locationd/cache_encryptedB.db : 0x34671E (Table: WifiLocation; Size: 4026368 bytes)
29.06.2023 12:13:02	Wireless Networks	BSSId: 44[redacted]	/private/var/root/Library/Caches/locationd/cache_encryptedB.db : 0x356886 (Table: WifiLocation; Size: 4026368 bytes)
29.06.2023 12:36:11	Wireless Networks	BSSId: B2[redacted]	/private/var/root/Library/Caches/locationd/cache_encryptedB.db : 0x36337D (Table: WifiLocation; Size: 4026368 bytes)
29.06.2023 12:57:31	Wireless Networks	BSSId: C4[redacted]	/private/var/root/Library/Caches/locationd/cache_encryptedB.db : 0x372343 (Table: WifiLocation; Size: 4026368 bytes)
29.06.2023 13:08:09	Wireless Networks	BSSId: 4A[redacted]	/private/var/root/Library/Caches/locationd/cache_encryptedB.db : 0x372318 (Table: WifiLocation; Size: 4026368 bytes)

Table 17: Harvested WiFi locations in Cellebrite



Tid	Aktivitet	Attributter	Database	Modul
29.06.2023 11:27:19	WiFi Location	TIMESTAMP: 29.06.2023 11:27:19 COORDINATES: 0.0, 0.0 MAC: 22[redacted] CHANNEL: -1 INFOMASK: 1 SPEED: -1.0 COURSE: -1.0 CONFIDENCE: 0 SCORE: -1 REACH: -1 HORIZONTAL ACCURACY: -1.0 VERTICAL ACCURACY: -1.0 LATITUDE: 0.0 LONGITUDE: 0.0	cache_encryptedB.db	locationd_cacheencryptedAB_wifilocation.txt#cache_encryptedB.db#SQL Query 8,9,10,11,12,13,10.13,10.14,10.15,10.16,14
29.06.2023 11:27:19	WiFi Location	TIMESTAMP: 29.06.2023 11:27:19 COORDINATES: 0.0, 0.0 MAC: 27[redacted] CHANNEL: -1 INFOMASK: 1 SPEED: -1.0 COURSE: -1.0 CONFIDENCE: 0 SCORE: -1 REACH: -1 HORIZONTAL ACCURACY: -1.0 VERTICAL ACCURACY: -1.0 LATITUDE: 0.0 LONGITUDE: 0.0	cache_encryptedB.db	locationd_cacheencryptedAB_wifilocation.txt#cache_encryptedB.db#SQL Query 8,9,10,11,12,13,10.13,10.14,10.15,10.16,14
29.06.2023 11:27:19	WiFi Location	TIMESTAMP: 29.06.2023 11:27:19 COORDINATES: 0.0, 0.0 MAC: 88[redacted] CHANNEL: -1 INFOMASK: 1 SPEED: -1.0 COURSE: -1.0 CONFIDENCE: 0 SCORE: -1 REACH: -1 HORIZONTAL ACCURACY: -1.0 VERTICAL ACCURACY: -1.0 LATITUDE: 0.0 LONGITUDE: 0.0	cache_encryptedB.db	locationd_cacheencryptedAB_wifilocation.txt#cache_encryptedB.db#SQL Query 8,9,10,11,12,13,10.13,10.14,10.15,10.16,14
29.06.2023 11:56:01	WiFi Location	TIMESTAMP: 29.06.2023 11:56:01 COORDINATES: 0.0, 0.0 MAC: 74[redacted] CHANNEL: -1 INFOMASK: 1 SPEED: -1.0 COURSE: -1.0 CONFIDENCE: 0 SCORE: -1 REACH: -1 HORIZONTAL ACCURACY: -1.0 VERTICAL ACCURACY: -1.0 LATITUDE: 0.0 LONGITUDE: 0.0	cache_encryptedB.db	locationd_cacheencryptedAB_wifilocation.txt#cache_encryptedB.db#SQL Query 8,9,10,11,12,13,10.13,10.14,10.15,10.16,14
29.06.2023 11:57:50	WiFi Location	TIMESTAMP: 29.06.2023 11:57:50 COORDINATES: 0.0, 0.0 MAC: 27[redacted] CHANNEL: -1 INFOMASK: 1 SPEED: -1.0 COURSE: -1.0 CONFIDENCE: 0 SCORE: -1 REACH: -1 HORIZONTAL ACCURACY: -1.0 VERTICAL ACCURACY: -1.0 LATITUDE: 0.0 LONGITUDE: 0.0	cache_encryptedB.db	locationd_cacheencryptedAB_wifilocation.txt#cache_encryptedB.db#SQL Query 8,9,10,11,12,13,10.13,10.14,10.15,10.16,14

Table 18: Harvested WiFi locations (without coordinates) in APOLLO

Summary
The assignments revealed several shortcomings in Cellebrite's presentation of details and contextual information. A lack of identifiers recursively affected the participants in the first example provided, making them unable to distinguish between connections to different Bluetooth devices. In the second example, a lack of contextual information led the participants to overlook the correct activity, instead relying on an erroneous activity. These findings reveal that clear and concise naming conventions, in conjunction with broad contextual information is important for the users ability to comprehend the meaning behind data.

6.1.3 Data aggregation

Data aggregation entails the compilation of data into summarized forms. In forensic tools, the approach could offer advantages such as ease of analysis and presentation, especially during the triage of large amounts of data. While similar, it differs from data normalization in that process intrinsically involves some form of data reduction, either by combining details or data points into summarized values.

Both tools in the study employ data normalization or aggregation and in one way or the other, either by combining data from different sources to unified categories or activities, or by aggregating data points into summary views. It can even be argued that the inherent purpose of these analysis tools is to aggregate and present source information in a streamlined fashion. Even though data aggregation can be discussed at several layers, this discussion is limited to aggregation as utilized by the tools in their presentation layers.

Several forms of data aggregation were identified throughout the assignments, with some forms affecting the participants more than others. While instances of aggregation were found in both tools, Cellebrite appears to implement such processes more frequently and to a larger than APOLLO. In order to clarify these instances, the section starts by mapping out different forms of aggregation, and discussed how they affected users in their various instances. These effects vary from introducing small misconceptions, to hinder the participants to answer any specifics regarding the information needs.

Hourly aggregation:
A form of data aggregation found unique to Cellebrite was the aggregation of data into hourly summaries. While the iOS operating system inherently performs aggregation on some data types, Cellebrite was found to expand the technique by aggregating granular data within the tool itself.

This practice was found to affect the Cellebrite group to a significant degree during their assessments of Task 8 (5.2.8). Among other things, the task concerned itself with the victim's last signs of life, an information need that depended on access to granular health data.

In instances of health data, Cellebrite was found to aggregate granular data into single summarizing records on an hourly basis. This summarization contains a mixture of health-related activities, with a single value representing each distinct activity type. For example: heart rate values were aggregated to a single value, containing the peak heart rate within the active period. Likewise, recorded steps were merged into a single total summarizing value. Not only did the practice obscure the chronological sequence of the data, but denied participants insight into data points that were essential to succeed in the assignment.

APOLLO, on the other side, showcased all health-related data, row by row, within their own distinct activity types. As opposed to Cellebrite, the group did not report any issues regarding the analysis of the devices health data.

The tools different approaches to data aggregation, resulted in a discrepancy between the performance of the groups. Regarding the victim's last signs of life, the Cellebrite group failed to assess any specifics, while the APOLLO group had no problems pinpointing the last signs of life in accordance with the ground truth (5.2.8).[9]

[9] The Cellebrite group was limited to an Excel spreadsheet exported through the tool's timeline view. Although the full version of Cellebrite Physical Analyzer presents aggregated health data as its main view, additional details can be viewed in its side pane. However, even though the tool offers additional details, it lacks options to export the data points. In addition, details can only be seen one hour at a time, within a mixture of several health data types. Even with the full version of the software, its practices make it difficult review any single data types such as heart rate samples over time.

Unified categories:
Both tools were found to compile data from similar sources into unified activities or categories. This process could potentially simplify a review, as users are relieved from navigating between multiple similar activities. For instance, if communications are scattered across several applications, unifying those conversations into a single category could facilitate a more coherent analysis of communications. On the other hand, unification of activities could lead to confusion, as users could struggle to differentiate between different sources of data.

Unification of data sources was found much more frequently in Cellebrite as apposed to APOLLO. In Cellebrite, a range of activities were unified under a limited set of categories. And as already discussed, the group had to differentiate between activities on the basis of somewhat ambiguous descriptors (6.1.1). Even though the same practice is implemented in APOLLO, it was only found in a limited set of activities.

The unification process appears to have affected the Cellebrite group positively in Task 1 (5.2.1). While the tool provided one single activity for all SMS communications, APOLLO segmented data from the SMS database into three unique activities; "SMS Chat", "SMS Chat - Message Delivered" and "SMS Chat - Message Read". Apart from one participant missing two messages, the group had no challenges identifying the communications. The APOLLO group on the other hand, redundantly selected several of the SMS-activities, showing multiple instances of the same massages. One participant in the group, only selected the "SMS Chat - Message Read" activity, leading them to only find received messages.

Periodic calculation:
Another aggregation form found unique to Cellebrite involves a periodic calculation of entities into single summarizing values. The logic is based upon the natural start and end times for distinct segments of data, and is implemented in instances where values signify change of statuses.

A concrete example of this was discovered during the groups assessment of Task 2 (5.2.2), where respondents were asked to establish connections to the victim's Wi-Fi network.

It was revealed that Cellebrite presents Wi-Fi channel usage from the database CurrentPowerlog.PLSQL in a radically different manner than APOLLO. While APOLLO presented granular data regarding the Wi-Fi channel usage, Cellebrite was found to perform periodic calculations of Wi-Fi channel usage, in order to determine the start and end time of distinct network connections. As a result, hundreds of rows containing Wi-Fi statuses were aggregated into single values signifying either "Connect" or "Disconnect".

While the approach implemented by Cellebrite seems more streamlined, none of the participants were able to identify the activity due to its ambiguous presentation (6.1.1). In addition, the same activity was misinterpreted as both Bluetooth connections and charging statuses in two of the other assignments.

Firstly, the activity was presented with ambiguous labels, stating either "Connect" or "Disconnect". Secondly, no additional information signified Wi-Fi channel usage or BSSIDs. Thirdly, hundreds of rows were aggregated into single statuses, signifying either "Connect" or "Disconnect". The combination of these factors clearly undermined the clarity and meaning of the activity.

Summary:
Three forms of data aggregation were identified throughout the assignments: hourly and categorical aggregations as well as aggregations of periodic calculations. While APOLLO occasionally parsed data from similar databases into unified activities, Cellebrite also performed post processing on data, either by summarizing values on an hourly basis, or by aggregating granular data into single rows that signified change of statuses.

Despite that the purpose of aggregation is to simplify presentation, it appeared more like a hindrance to the participants as important contextual information was removed. The gravest example of this could be seen in Cellebrite's hourly aggregation of health data, as the participants were unable to provide any specifics about the data.

While the compilation of data into presentable formats can make data analysis more efficient, it seems that a certain threshold of information loss leads to diminishing returns. As the act of data normalization intrinsically entails some form of data reduction, there is the potential that information that is vital for an investigation could get lost.

6.1.4 Timestamps

Interpretation of timestamps is a vital part of most digital forensic investigations, and holds significant importance when establishing sequences of events. Determining the validity of timestamps could be associated with a plethora of challenges, such as their accuracy and reliability, time zone, formatting, and in certain instances, calculation of their offset values.

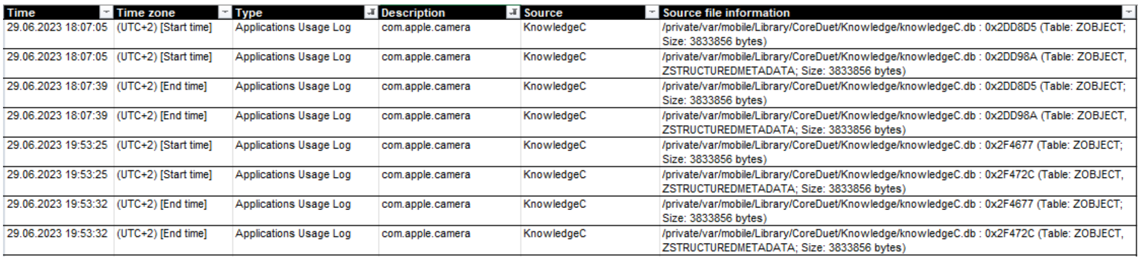
Both Cellebrite and APOLLO facilitate timestamp normalization through their presentation layers, implying that formats such MAC absolute time and UNIX time are converted into a human readable format. Despite that both tools convey normalized outputs, variations in their presentations were found to affect the assignments.

A key difference between the tools was found in the presentation of artifacts with intrinsic time frames. A concrete example of this is presented in Table 19 and 20, showcasing time frames for usage of the device's camera application in Cellebrite and APOLLO.

As seen in Table 19, Cellebrite presented redundant start and end timestamps, as application usage is seemingly parsed twice. The tool also segregated start and end times into two separate rows, despite that the data is sourced from KnowledgeC.db, where the start and end time is stored within a single record.

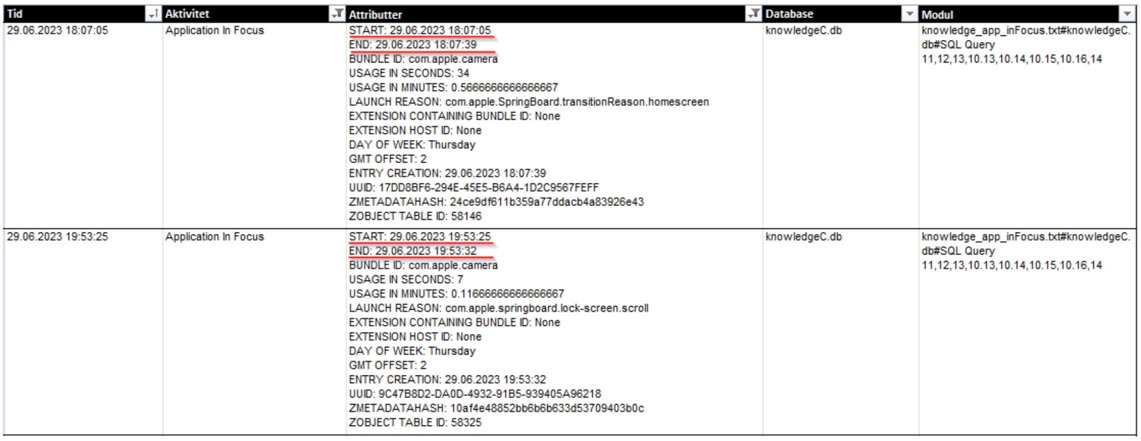
The same information presented through APOLLO can be seen in Table 20. As opposed to Cellebrite's eight rows of information, the same data was presented with only two rows in APOLLO. The tool also retained the originality of the data by preserving the start and end time within a single row. While APOLLO occasionally had two modules, one for the start and one for end, both rows contained the same information signifying the duration of the event.

Throughout the assessments, it became evident that the Cellebrite group experienced difficulties when attempting to determine the duration of events. The issue seems to be worsened when evaluating several activity types at once, and when records from different activities were intermixed. In these instances, the participants were forced to search the results for matching start and end records, experienced as a tedious process by the participants. The APOLLO group, on the other hand, did not experience the same degree of confusion, first and foremost due to time frames being contained within single rows.



Time	Time zone	Type	Description	Source	Source file information
29.06.2023 18:07:05	(UTC+2) [Start time]	Applications Usage Log	com.apple.camera	KnowledgeC	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x2DD8D5 (Table: ZOBJECT; Size: 3833856 bytes)
29.06.2023 18:07:05	(UTC+2) [Start time]	Applications Usage Log	com.apple.camera	KnowledgeC	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x2DD98A (Table: ZOBJECT, ZSTRUCTUREDMETADATA; Size: 3833856 bytes)
29.06.2023 18:07:39	(UTC+2) [End time]	Applications Usage Log	com.apple.camera	KnowledgeC	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x2DD8D5 (Table: ZOBJECT; Size: 3833856 bytes)
29.06.2023 18:07:39	(UTC+2) [End time]	Applications Usage Log	com.apple.camera	KnowledgeC	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x2DD98A (Table: ZOBJECT, ZSTRUCTUREDMETADATA; Size: 3833856 bytes)
29.06.2023 19:53:25	(UTC+2) [Start time]	Applications Usage Log	com.apple.camera	KnowledgeC	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x2F4677 (Table: ZOBJECT; Size: 3833856 bytes)
29.06.2023 19:53:25	(UTC+2) [Start time]	Applications Usage Log	com.apple.camera	KnowledgeC	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x2F472C (Table: ZOBJECT, ZSTRUCTUREDMETADATA; Size: 3833856 bytes)
29.06.2023 19:53:32	(UTC+2) [End time]	Applications Usage Log	com.apple.camera	KnowledgeC	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x2F4677 (Table: ZOBJECT; Size: 3833856 bytes)
29.06.2023 19:53:32	(UTC+2) [End time]	Applications Usage Log	com.apple.camera	KnowledgeC	/private/var/mobile/Library/CoreDuet/Knowledge/knowledgeC.db : 0x2F472C (Table: ZOBJECT, ZSTRUCTUREDMETADATA; Size: 3833856 bytes)

Table 19: Camera usage in Cellebrite



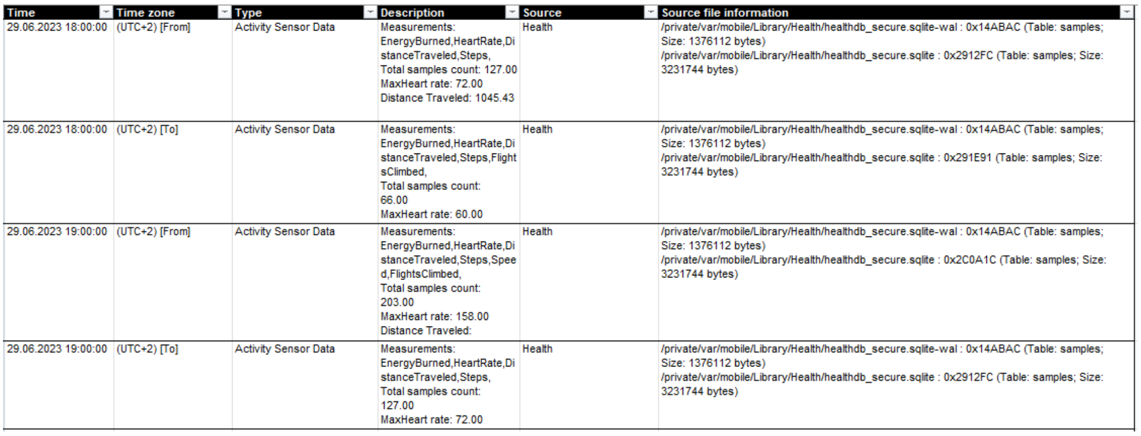
Tid	Aktivitet	Attributter	Database	Modul
29.06.2023 18:07:05	Application In Focus	START: 29.06.2023 18:07:05 END: 29.06.2023 18:07:39 BUNDLE ID: com.apple.camera USAGE IN SECONDS: 34 USAGE IN MINUTES: 0.5666666666666667 LAUNCH REASON: com.apple.SpringBoard.transitionReason.homescreen EXTENSION CONTAINING BUNDLE ID: None EXTENSION HOST ID: None DAY OF WEEK: Thursday GMT OFFSET: 2 ENTRY CREATION: 29.06.2023 18:07:39 UUID: 17DD8BF6-294E-45E5-B6A4-1D2C9567FEFF ZMETADATAHASH: 24ce9df611b359a77ddacb4a83926e43 ZOBJECT TABLE ID: 58146	knowledgeC.db	knowledge_app_inFocus.txt#knowledgeC.db#SQL Query 11,12,13,10.13,10.14,10.15,10.16,14
29.06.2023 19:53:25	Application In Focus	START: 29.06.2023 19:53:25 END: 29.06.2023 19:53:32 BUNDLE ID: com.apple.camera USAGE IN SECONDS: 7 USAGE IN MINUTES: 0.11666666666666667 LAUNCH REASON: com.apple.springboard.lock-screen.scroll EXTENSION CONTAINING BUNDLE ID: None EXTENSION HOST ID: None DAY OF WEEK: Thursday GMT OFFSET: 2 ENTRY CREATION: 29.06.2023 19:53:32 UUID: 9C47B8D2-DA0D-4932-91B5-939405A96218 ZMETADATAHASH: 10af4e48852bb6b6b633d53709403b0c ZOBJECT TABLE ID: 58325	knowledgeC.db	knowledge_app_inFocus.txt#knowledgeC.db#SQL Query 11,12,13,10.13,10.14,10.15,10.16,14

Table 20: Camera usage in APOLLO

The Cellebrite group experienced additional challenges when assessing Task 8 (5.2.8). As discussed earlier, the tool was found to aggregate health data into hourly summaries, a process that denied them access to granular health data. In relation to these hourly summaries, respondents experienced uncertainty when determining the actual start and end times of each aggregated time period, due to the way they were presented.

In the post-scenario interviews, one participant experienced confusion when analyzing activities that had identical start and end times (5.3.4), while another explained that it was difficult to establish the actual start and end times of distinct events, as had to scan the table in an attempt to match start rows with corresponding end rows.

The reason behind the confusion are illustrated in Table 21, showcasing how the tool produced two records each hour, each signifying the start or end of an aggregated period. As the tool's timeline sorts data based on its timestamps, the natural ordering of each hour's start and end periods was presented randomly, leading to confusion whether a time period had started or ended.



Time	Time zone	Type	Description	Source	Source file information
29.06.2023 18:00:00	(UTC+2) [From]	Activity Sensor Data	Measurements: EnergyBurned,HeartRate,DistanceTraveled,Steps, Total samples count: 127.00 MaxHeart rate: 72.00 Distance Traveled: 1045.43	Health	/private/var/mobile/Library/Health/healthdb_secure.sqlite-wal : 0x14ABAC (Table: samples; Size: 1376112 bytes) /private/var/mobile/Library/Health/healthdb_secure.sqlite : 0x2912FC (Table: samples; Size: 3231744 bytes)
29.06.2023 18:00:00	(UTC+2) [To]	Activity Sensor Data	Measurements: EnergyBurned,HeartRate,DistanceTraveled,Steps,FlightsClimbed, Total samples count: 66.00 MaxHeart rate: 60.00	Health	/private/var/mobile/Library/Health/healthdb_secure.sqlite-wal : 0x14ABAC (Table: samples; Size: 1376112 bytes) /private/var/mobile/Library/Health/healthdb_secure.sqlite : 0x291E91 (Table: samples; Size: 3231744 bytes)
29.06.2023 19:00:00	(UTC+2) [From]	Activity Sensor Data	Measurements: EnergyBurned,HeartRate,DistanceTraveled,Steps,Speed,FlightsClimbed, Total samples count: 203.00 MaxHeart rate: 158.00 Distance Traveled:	Health	/private/var/mobile/Library/Health/healthdb_secure.sqlite-wal : 0x14ABAC (Table: samples; Size: 1376112 bytes) /private/var/mobile/Library/Health/healthdb_secure.sqlite : 0x2C0A1C (Table: samples; Size: 3231744 bytes)
29.06.2023 19:00:00	(UTC+2) [To]	Activity Sensor Data	Measurements: EnergyBurned,HeartRate,DistanceTraveled,Steps, Total samples count: 127.00 MaxHeart rate: 72.00	Health	/private/var/mobile/Library/Health/healthdb_secure.sqlite-wal : 0x14ABAC (Table: samples; Size: 1376112 bytes) /private/var/mobile/Library/Health/healthdb_secure.sqlite : 0x2912FC (Table: samples; Size: 3231744 bytes)

Table 21: Aggregated health data in Cellebrite

Summary:
The tools were found to convey distinct characteristics regarding their presentation of timestamps. While both tools normalized timestamps to a human readable format, Cellebrite was found to perform additional post-processing on the timestamps.

The manner in which the tools presented their timestamps were found to have a significant impact on the participants' ability to identify correct time frames. Overall, the participants experienced the less processed results as easier to interpret.

6.2 Group performance

In the preceding section, distinct patterns within the tools presentation layers were synthesized into broad categories. These categories include weak and ambiguous naming conventions, data aggregation, lack of details and contextual information, and the tools presentation of timestamps. Each category was found to impact participant assessments, either by obscuring the respondents ability to interpret data or by denying them access to essential details.

Viewed separately, each category affected a specific set of assignments. For example, weak and ambiguous naming conventions led to misinterpretation of activities in certain instances, and data aggregation hindered access to essential details in others. Despite that each individual element exerted an impact on its own, some assessments were affected by several concurrent elements. The extent to which each group was affected differed, in that the cumulative effect of all categories across the assignments led to notable differences between the groups.

6.2.1 Where did the groups perform on par?

The groups performance was tightly related to the complexity and details required for each task. In basic tasks, such as providing SMS communications, browsing history, and activation of airplane mode (5.2.1, 5.2.5, 5.2.9), the performance levels between the groups were similar. Common factors between these tasks, were that both tools provided clear naming conventions for relevant activities, and that none of the assignments required the participants to establish time frames for the events.

While both groups assessed the above mentioned tasks with relative ease, some minor misconceptions emerged. In one instance, a participant in the APOLLO group only provided incoming SMS message due to the tool's disparity of SMS activities. In this particular instance, the participant focused exclusively on SMS read confirmations, a factor that could be attributed to the tools naming conventions. In another instance, a participant in the Cellebrite group failed to provide all relevant SMS communications due to neglect.

6.2.2 The essence of differences

While the groups performance levels was on par in the most basic tasks, it became evident that the discrepancy increased with the intricacy of the assignments. The remaining tasks introduced additional levels of complexity that included the need to establish sequences of events, and to determine distinct time frames for activities. Additionally, several assignments required specific details, such as identifiers that could be linked to external devices.

Cellebrite's tendency to unify activities into broad categories, paired with its somewhat ambiguous naming conventions led the respondents to misinterpret several activities. This effect was amplified as the tool often lacked contextual information that otherwise would have helped infer meaning to its activities. Despite that the group was able to assess several of the assignments, the shortcomings often resulted in vague and imprecise answers.

A highly problematic factor was Cellebrite's tendency to dividing naturally occurring time frames into separate rows. This practice was found to impede the group's ability to efficiently assess the duration of events, as the group was forced to scan the results for corresponding start and end rows. Despite that the cumbersome process yielded some results, the respondents experienced an overarching confusion related to activities with time frames.

At the core of the confusion was the tool's tendency to present the start of a new period with identical timestamps to the end of a previous period. The confusion was amplified as the tool sorts results on timestamps, resulting in start and end periods (with identical timestamps) presented in an arbitrary order. This phenomenon made the participants uncertain whether distinct time periods actually had started or ended.

APOLLO displayed significantly different characteristics when it comes to the above mentioned elements. The tool offers a much more granular view of its data, providing a total of 55,222 rows from 16 databases, compared to 13,979 rows from 12 databases in the case of Cellebrite (3.10). APOLLO also offered greater segmentation of its data, providing a total of 101 unique activities within the time period of the scenario. Cellebrite, on the other hand, had a tendency to unify its activities into broad and generic categories, leading the participants to infer meaning based on its somewhat generic labels. The detail level for each distinct artifact also surpassed that of Cellebrite, allowing the group to infer meaning to its activities based on the contextual richness, a factor that reduced the group number of misinterpretations.

Overall APOLLO provided a less processed view of the source data compared to Cellebrite. Despite that the post-processing performed by Cellebrite most likely is aimed to enhance the user experience, it was shown to be counterproductive for the groups performance, especially when answering the more intricate assessments.

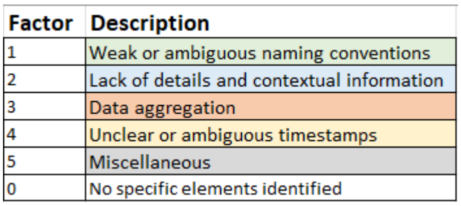
6.2.3 Measurement of performance

Although the qualitative aspects of the groups performance have been discussed in detail, a quantitative metric could shed further light on the research question. However, to achieve this, a concrete numerical performance measure for each assessment would have been needed.

Given the limited number of respondents, varying depth of responses, and the complexity of certain assignments, a mixed methods approach seemed more appropriate. Rather than relying on a qualitative evaluation alone, quantitative insights of the groups performance can be inferred from the preceding analysis and discussions. By counting occurrences of the categories synthesized above, each task can be given an approximate error rate.

The synthesized categories were either found to hinder the groups ability to answer distinct information needs, or lead to some form of confusion or misinterpretation. The categories include weak and ambiguous naming conventions, lack of details and contextual information, data aggregation, unclear or ambiguous timestamps, and other miscellaneous elements.

For purposes of enumeration and visualization, each synthesized element was given a unique number and color code as illustrated in Table 22. These markers are used consistently throughout the figures that follow.



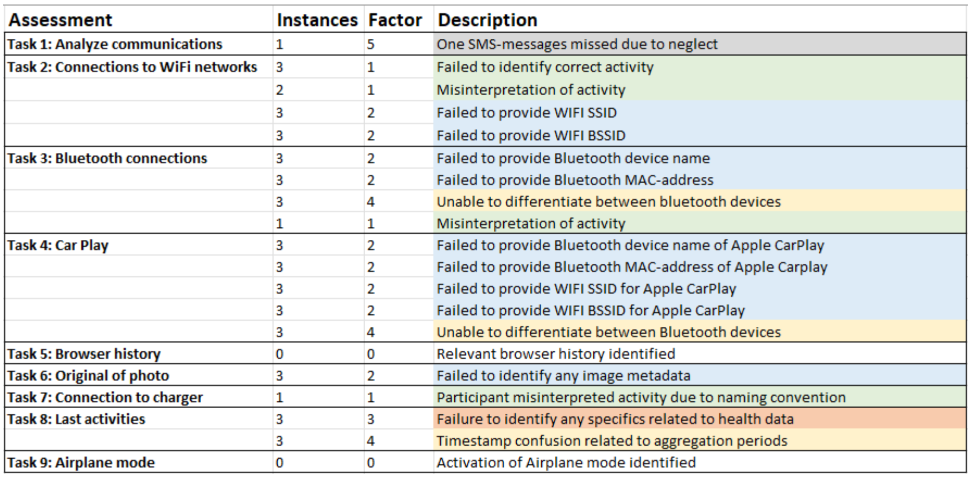
Factor	Description
1	Weak or ambiguous naming conventions
2	Lack of details and contextual information
3	Data aggregation
4	Unclear or ambiguous timestamps
5	Miscellaneous
0	No specific elements identified

Table 22: Synthesized elements in the tools presentation layers

Cellebrite group:
The analysis and discussion revealed a total of 46 failures that can be attributed to the synthesized categories. Table 23 provides an overview of these instances and how they are distributed among the elements. Figure 21 provides a bar chart of the data, visualizing the distribution of errors.

The distribution of these errors was as follows:

- 28 of the instances were attributed to a lack of detail or contextual information.
- 8 of the instances were associated with timestamp confusions.
- 7 of the instances were related to weak and ambiguous naming conventions.
- 3 of the instances were related to data aggregation.
- 1 of the instances could be attributed to other miscellaneous elements.



Assessment	Instances	Factor	Description
Task 1: Analyze communications	1	5	One SMS-messages missed due to neglect
Task 2: Connections to WiFi networks	3	1	Failed to identify correct activity
	2	1	Misinterpretation of activity
	3	2	Failed to provide WIFI SSID
	3	2	Failed to provide WIFI BSSID
Task 3: Bluetooth connections	3	2	Failed to provide Bluetooth device name
	3	2	Failed to provide Bluetooth MAC-address
	3	4	Unable to differentiate between bluetooth devices
	1	1	Misinterpretation of activity
Task 4: Car Play	3	2	Failed to provide Bluetooth device name of Apple CarPlay
	3	2	Failed to provide Bluetooth MAC-address of Apple Carplay
	3	2	Failed to provide WIFI SSID for Apple CarPlay
	3	2	Failed to provide WIFI BSSID for Apple CarPlay
	3	4	Unable to differentiate between Bluetooth devices
Task 5: Browser history	0	0	Relevant browser history identified
Task 6: Original of photo	3	2	Failed to identify any image metadata
Task 7: Connection to charger	1	1	Participant misinterpreted activity due to naming convention
Task 8: Last activities	3	3	Failure to identify any specifics related to health data
	3	4	Timestamp confusion related to aggregation periods
Task 9: Airplane mode	0	0	Activation of Airplane mode identified

Table 23: Errors encountered in the Cellebrite group

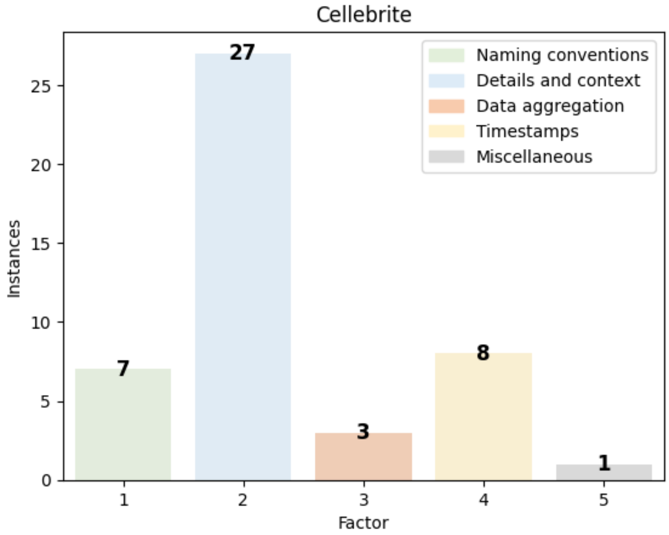
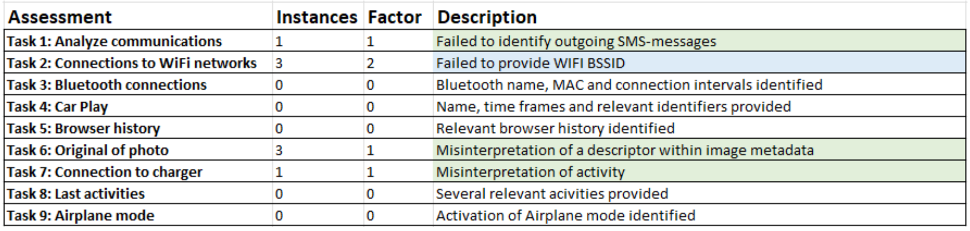


Figure 21: Distribution of errors in the Cellebrite group

APOLLO group:
A total of 8 errors were found in the APOLLO group that can be attributed to the synthesized categories. Table 24 provides an overview of these instances and how they are distributed among the categories. Figure 22 provides a bar chart of the data, visualizing the distribution of errors.

The APOLLO group exerted a total of 8 failures that could be attributed to the tool's presentation layer.

- 6 instances were attributed to details and contextual information.
- 2 instances were attributed to its naming conventions.



Assessment	Instances	Factor	Description
Task 1: Analyze communications	1	1	Failed to identify outgoing SMS-messages
Task 2: Connections to WiFi networks	3	2	Failed to provide WIFI BSSID
Task 3: Bluetooth connections	0	0	Bluetooth name, MAC and connection intervals identified
Task 4: Car Play	0	0	Name, time frames and relevant identifiers provided
Task 5: Browser history	0	0	Relevant browser history identified
Task 6: Original of photo	3	1	Misinterpretation of a descriptor within image metadata
Task 7: Connection to charger	1	1	Misinterpretation of activity
Task 8: Last activities	0	0	Several relevant acivities provided
Task 9: Airplane mode	0	0	Activation of Airplane mode identified

Table 24: Errors encountered in the APOLLO group

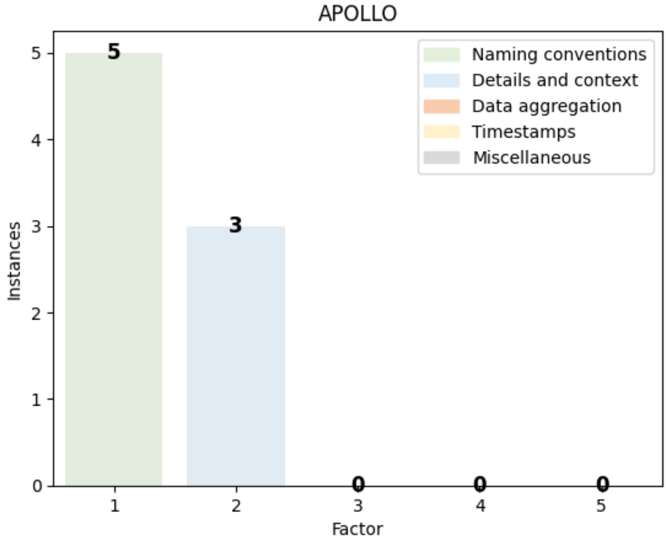


Figure 22: Distribution of instances in the APOLLO group

6.2.4 Conclusion of the performance measure

It is important to emphasize that the error rates were inferred from the qualitative analysis of the participant responses. As such, they are approximate estimates used to showcase the distribution of errors between the synthesized categories. Despite the potential for some missing or wrongfully categorized instances, the error rates are believed to be a valuable addition to the qualitative analysis.

The error rates illustrate a major discrepancy between the performance of the groups, as a total of 46 errors were found in the Cellebrite group compared to 8 in the Apollo group.

Lack of details imposed a significant disadvantage for the Cellebrite group, leading to a cumulative effect of 27 errors, including lack of specifics and misinterpretation of activities. APOLLO, with its more granular details, resulted in a total of 3 errors.

Misinterpretations of timestamps resulted in 8 errors in the Cellebrite group. 2 of the instances coincided with lacking details, 3 followed recursively from the first instance, and 3 coincided with aggregation of health data. No errors were found in the APOLLO group.

The third most frequent factor was weak and ambiguous naming conventions. The Cellebrite group produced a total of 7 errors, whereby 3 instances resulted from failure to identify the correct activity, and 4 resulted from misinterpretation of activities. The APOLLO group produced 5 errors, 1 due to granularity of activities and 4 due to misinterpretation of an activities.

Aggregation of health data led to 3 errors in the Cellebrite group, all resulting in a failure to produce specific details. The APOLLO group was not affected by data aggregation to a degree that affected the assignments.

The quantitative analysis highlights lack of details as the most prominent factor between the groups, followed by timestamps, naming conventions and data aggregation. The differences in performance between the groups underscores that the way in which digital forensic tools present their data can have a significant impact on investigations.

7 Conclusion

The research questions of the thesis aimed to investigate how factors within the presentation layer of digital forensic analysis tools affect investigator opinions, and to what extent they affect criminal investigations. The study identified a previously unknown source of bias in digital forensics, resulting in a classification of *data presentation biases in forensic analysis.*

What factors within the presentation layer of digital forensic analysis tools affect investigator opinions?

The study synthesized four main areas within forensic tool presentation layers found to affect investigator opinions:

- Naming conventions
- Details and contextual information
- Data aggregation
- Timestamps

In some instances, single factors such as naming conventions or lack of details affected assessments independently. In other instances, several factors impacted an assessment simultaneously, leading to a compounding effect.

The participants displayed a tendency to interpret the tools activity names literally. As a consequence, the error rate was higher in instances where the tool provided ambiguous names for its activities and descriptors, and where limited contextual information was available. Activities were misinterpreted in two main ways: firstly, through a literal interpreting of the tools descriptors, and secondly, due to challenges in differentiating between events within distinct activities. Misinterpretations due to ambiguous naming conventions appear to be closely related to insufficient details, as participants in many cases would have recognized their true meaning with sufficient contextual information.

Respondents were also affected by the practice of data aggregation. In several instances, data aggregation suppressed the granularity of data to such a degree, that they were unable to answer any specifics. This was especially true regarding Cellebrite's presentation of health data, where several data types were merged into hourly summaries.

The way in which the tools presented timestamps was also a factor for misinterpretation. The participants struggled to establish the duration of events in instances where start and end times were separated into distinct rows. This practice necessitated the participants to search through the results for matching start and end rows, usually with multiple rows of data in between. The combination of aggregated data with separated time frames made it even more confusing, leading participants to be uncertain whether distinct time periods had started or ended.

Lack of details and contextual information contribute to errors both individually and in combination with other factors. In some cases, lack of detail led to confusion regarding the meaning of various data points, in other instances, it hindered the participants to answer concrete information needs.

If such factors are found, to what extent do they affect criminal investigations?

The final results showcased a major discrepancy between the error rates of each group, with a total of 46 instances in the Cellebrite group, compared to seven instances in the APOLLO group.

Data for the scenario was generated through real-life actions to simulate a suspicious death. The participants were then tasked with answer information needs relevant to the initial phase of the investigation. Each information need was constructed to answer pieces of the story that led up to the victim's death. Not only did answering these information needs give context to the day of the victim and the time of her death, but gave information actionable to identifying the perpetrator. As it follows, each failed attempt in answering these information needs was a missed opportunity to uncover pieces of the ground truth.

The discrepancy between the groups was clearly related to the presentation layer of each tool, and illustrates how design choices such as data normalization and post processing of results can impair investigators' ability to answer relevant information needs in criminal investigations.

A classification of data presentation biases in forensic analysis:

Itiel E Dror (2020) demonstrates how contextual information, without case-specific relevance, can lead to biases that affect how experts view, interpret, and conclude their findings. This specific factor is part of a broader framework that encompasses seven additional biases, all influencing forensic experts in their decision-making processes.

This study has expanded upon a pre-existing knowledge-base by showcasing how the presentation layer in digital forensic tools also can mislead experts by introducing biases and misinterpretations. In light of existing research on biases within digital forensics, findings of this study point towards a new category of biases that has not previously been examined. Therefore, the study introduces the classification *data presentation biases in forensic analysis* as a way to address this specific category of biases.

The study serves as a preliminary discussion into these findings, integrating a limited number of categories. Due to limitations of the study, the data is not sufficient to make any definite statements regarding the consistency of the misinterpretations that were found. It does however rise questions whether the tool's presentation of data also could lead so systematic biases in interpretations.

These aspects call for a further refinement of the categories that were discovered, and an expansion into additional factors and their effects. It highlights the need to develop strategies for mitigation that seek to minimize the impact of tool presentation biases in the future.

References

Braun, Virginia and Victoria Clarke (2006). Using thematic analysis in psychology. eng. In: Qualitative research in psychology 3.2, pp. 77–101. ISSN: 1478-0887.

Brignoni, Alexis (2023). iOS Logs, Events, And Properties Parser (iLEAPP). URL: https://github.com/abrignoni/iLEAPP. Accessed: 2023-12-12.

Carrier, Brian (2002). Open source digital forensics tools: The legal argument.

Casey, Eoghan (2002). Error, uncertainty and loss in digital evidence. In: International Journal of Digital Evidence 1.2.

— (2004). Digital Evidence and Computer Crime: Forensic Science, Computers and the Internet. eng. Academic Press. ISBN: 0121631044.

— (2019). The chequered past and risky future of digital forensics. eng. In: Australian journal of forensic sciences 51.6, pp. 649–664. ISSN: 0045-0618.

Cellebrite (2023). Cellebrite Physical Analyzer. URL: https://cellebrite.com/en/physical-analyzer. Accessed: 2023-12-12.

Cooper, Glinda S. and Vanessa Meterko (2019). Cognitive bias research in forensic science: A systematic review. eng. In: Forensic science international 297, pp. 35–46. ISSN: 0379-0738.

Craddock, Rachel, Doug Watson and William Saunders (2016). Generic Pattern of Life and behaviour analysis. IEEE.

Dror, I. E. and J. L. Mnookin (2010). The use of technology in human expert domains: challenges and risks arising from the use of automated fingerprint identification systems in forensic science. eng. In: Law, probability and risk 9.1, pp. 47–67. ISSN: 1470-8396.

Dror, Itiel E (2020). Cognitive and Human Factors in Expert Decision Making: Six Fallacies and the Eight Sources of Bias. eng. In: Analytical chemistry (Washington) 92.12, pp. 7998–8004. ISSN: 0003-2700.

Dror, Itiel E. et al. (2012). The Impact of Human-Technology Cooperation and Distributed Cognition in Forensic Science: Biasing Effects of AFIS Contextual Information on Human Experts. eng. In: Journal of forensic sciences 57.2, pp. 343–352. ISSN: 0022-1198.

Edwards, Sarah (2019). Launching APOLLO. URL: https://github.com/mac4n6/Presentations/blob/master/LaunchingAPOLLO/LaunchingAPOLLO.pdf. MacDevOpsYVR 2019. URL: https://github.com/mac4n6/Presentations/blob/master/LaunchingAPOLLO/LaunchingAPOLLO.pdf.

— (2020a). Apple Pattern of Life Lazy Output'er (APOLLO). URL: https://github.com/mac4n6/apollo. Accessed: 2023-12-12.

— (2020b). Getting Spooky with APOLLO. URL: https://github.com/mac4n6/Presentations/blob/master/Getting%20Spooky%20with%20APOLLO/GettingSpookyWithAPOLLO.pdf.

— (2020c). Go For Launch - Getting Started with Pratical APOLLO Analysis. URL: https://github.com/mac4n6/Presentations/blob/master/Go%20For%20Launch%20-%20Getting%20Started%20with%20Pratical%20APOLLO%20Analysis/GoForLaunch.pdf.

Erlandsen, Tom Erik (2019). Fallacies when Evaluating Digital Evidence Among Prosecutors in the Norwegian Police Service. eng. URL: http://hdl.handle.net/11250/2617771.

Fukami, A, R Stoykova and Z Geradts (2021). A new model for forensic data extraction from encrypted mobile devices. eng. In: Forensic Science International: Digital Investigation 38. ISSN: 2666-2817.

Henseler, Hans (2020). The Potential of Digital Traces in Providing Evidence at Activity Level. In: Proceedings of the Digital Forensic Research Workshop. URL: https://dfrws.org/wp-content/uploads/2020/06/2020_USA_pres-the_potential_of_digital_traces_in_providing_evidence_at_activity_level-1.pdf.

Hensler, Hans (2022). The Potential of Digital Traces in Providing Evidence at Activity Level. eng. In: Forensic Focus.

Hickman, Joshua (2021). iOS 14 & macOS Big Sur: Lots of Images. URL: https://thebinaryhick.blog/2021/02/20/ios-14-macos-big-sur-lots-of-images. Accessed on: 2023-11-12.

Horsman, Graeme (2022). "Scaffolding" responses to digital forensic inquiries. eng. In: WIREs. Forensic science 4.4, e1451–n/a. ISSN: 2573-9468.

Ibbetson, Ross (2023). Read minute-by-minute breakdown of Maggie and Paul Murdaugh's cell phone activity on night they were executed: Everything from calls, texts, and videos - to even their FOOTSTEPS. Accessed: 2023-05-24. URL: https://www.dailymail.co.uk/news/article-

11703281/Minute-minute-breakdown-Maggie-Paul-Murdaughs-cell-phone-movements-night-died.html.

LB-2022-17642. Borgating lagmannsrett (2022). URL: https://lovdata.no/dokument/LBSTR/avgjorelse/lb-2022-17642.

Leedy, Paul D (2021). Practical research : planning and design. eng. Harlow.

McQuade, Sam (2006). Technology-enabled Crime, Policing and Security. eng. In: The Journal of technology studies 32.1/2, pp. 32–42. ISSN: 1071-6084.

Meconi, Timo and Hans Henseler (2022). Digitale sporen in smartphones. Een kennismaking met pattern-of-life forensics. In.

NPCC, National Police Chiefs Council (2020). Digital Forensic Science Strategy. URL: https://www.npcc.police.uk/SysSiteAssets/media/downloads/publications/publications-log/2020/national-digital-forensic-science-strategy.pdf (visited on 1st Feb. 2021).

Pollitt, Mark M (1995). Computer forensics: An approach to evidence in cyberspace. In: Proceedings of the 18th National Information Systems Security Conference, pp. 487–491.

Rudofski, Special Agent Peter (2023). Murdaugh Murders Timeline of Events: Condensed Timeline. Document provided by user. Accessed: 2023-11-12. URL: https://www.counton2.com/wp-content/uploads/sites/7/2023/02/Condensed-Timeline-1.pdf.

Smit, Nadine M., Ruth M. Morgan and David A. Lagnado (2018). A systematic analysis of misleading evidence in unsafe rulings in England and Wales. eng. In: Science & justice 58.2, pp. 128–137. ISSN: 1355-0306.

Sunde, Nina (2021). What does a digital forensics opinion look like? A comparative study of digital forensics and forensic science reporting practices. eng. In: Science & justice 61.5, pp. 586–596. ISSN: 1355-0306.

— (2022). Unpacking the evidence elasticity of digital traces. eng. In: Cogent social sciences 8.1. ISSN: 2331-1886.

Sunde, Nina and Itiel E. Dror (2021). A hierarchy of expert performance (HEP) applied to digital forensics: Reliability and biasability in digital forensics decision making. eng. In: Forensic Science International: Digital Investigation 37, p. 301175. ISSN: 2666-2817.

TOSL-2022-14854. Oslo tingrett (2022). URL: https://lovdata.no/dokument/TRSTR/avgjorelse/tosl-2022-14854.

TOSLO-2020-20518-2. Oslo tingrett (2020). URL: https://lovdata.no/dokument/TRSTR/avgjorelse/toslo-2020-20518-2.

Whiffin, Ian (2022). Harvested Locations. Accessed: 2023-08-05. URL: https://www.doubleblak.com/blogPosts.php?id=16.

Williams, Janet (2012). Acpo good practice guide for digital evidence. In: Metropolitan Police Service, Association of chief police officers, GB, pp. 1556–6013.

Wilson-Kovacs, Dana et al. (2023). Digital evidence in defence practice: Prevalence, challenges and expertise. eng. In: The international journal of evidence & proof, p. 136571272311716. ISSN: 1365-7127.

WLTX, News 19 (2023). Alex Murdaugh detailed timeline revealed by cell phone and car data: full video. Accessed: 2023-05-24. URL: https://www.youtube.com/watch?v=EU0NBRej6D4.

Yeboah-Ofori, Abel et al. (2019). Relativism Digital Forensics Investigations Model: A Case for the Emerging Economies. 2019 International Conference on Cyber Security and Internet of Things (ICSIoT). Piscataway, NJ : IEEE, pp. 14–100. ISBN: 1-7281-7418-X.

You, Ilsun et al. (2022). Forensic Analysis of Apple CarPlay: A Case Study. eng. In: vol. 1544. Communications in Computer and Information Science. Singapore: Springer Singapore Pte. Limited, pp. 289–300. ISBN: 981169575X.

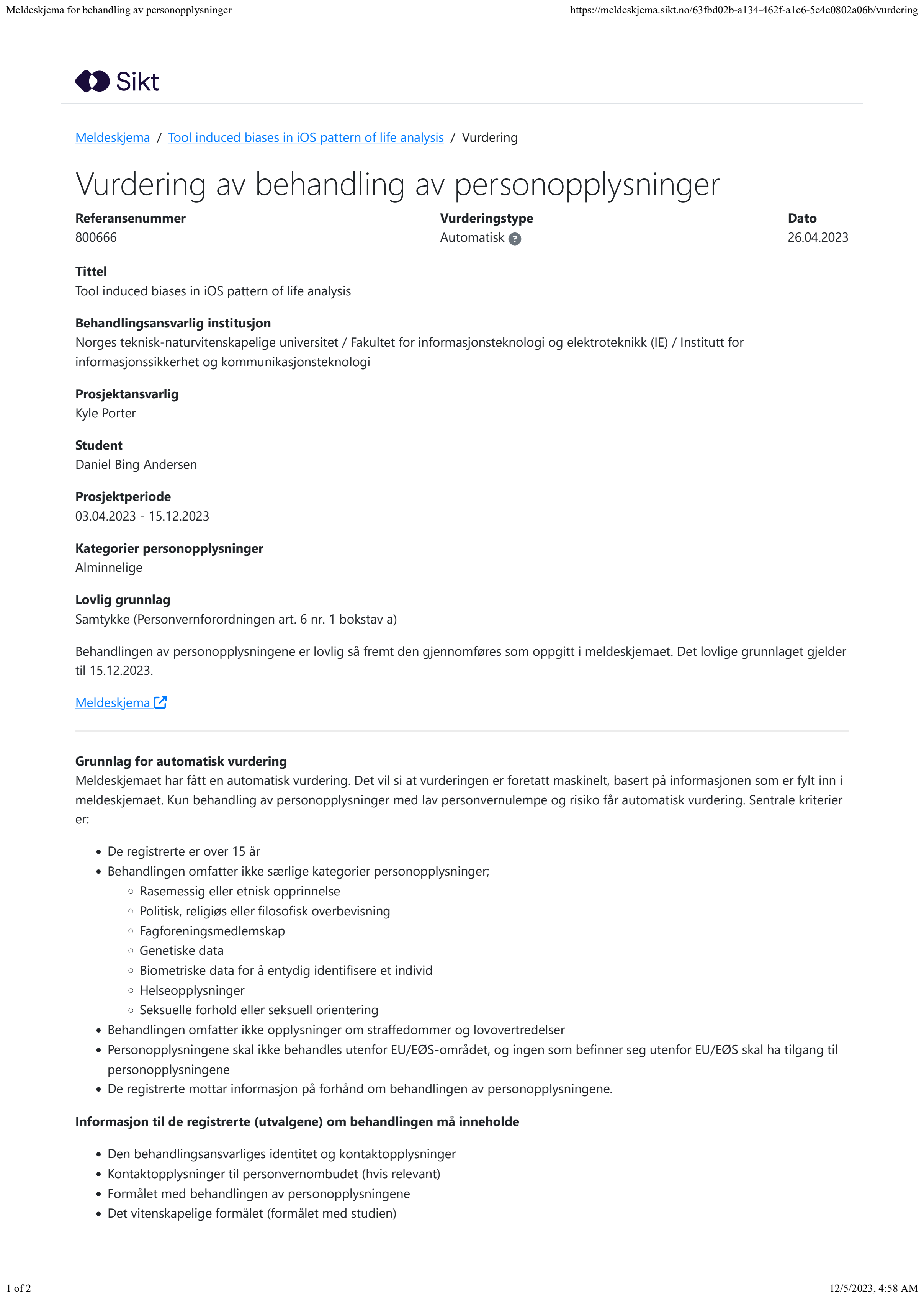
Zandwijk, Jan Peter van and Abdul Boztas (2021). The phone reveals your motion: Digital traces of walking, driving and other movements on iPhones. eng. In: Forensic Science International: Digital Investigation 37, p. 301170. ISSN: 2666-2817.

— (2019). The iPhone Health App from a forensic perspective: can steps and distances registered during walking and running be used as digital evidence? eng. In: Digital investigation 28, S126–S133. ISSN: 1742-2876.

Appendices

Appendix 1: Sikt approval
Appendix 2: Consent form
Appendix 3: Question form
Appendix 4: Interview guide
Appendix 5: Activity types APOLLO

1 Sikt approval



Sikt

Meldeskjema / Tool induced biases in iOS pattern of life analysis / Vurdering

Vurdering av behandling av personopplysninger

Referansenummer	Vurderingstype	Dato
800666	Automatisk ?	26.04.2023

Tittel
Tool induced biases in iOS pattern of life analysis

Behandlingsansvarlig institusjon
Norges teknisk-naturvitenskapelige universitet / Fakultet for informasjonsteknologi og elektroteknikk (IE) / Institutt for informasjonssikkerhet og kommunikasjonsteknologi

Prosjektansvarlig
Kyle Porter

Student
Daniel Bing Andersen

Prosjektperiode
03.04.2023 - 15.12.2023

Kategorier personopplysninger
Alminnelige

Lovlig grunnlag
Samtykke (Personvernforordningen art. 6 nr. 1 bokstav a)

Behandlingen av personopplysningene er lovlig så fremt den gjennomføres som oppgitt i meldeskjemaet. Det lovlige grunnlaget gjelder til 15.12.2023.

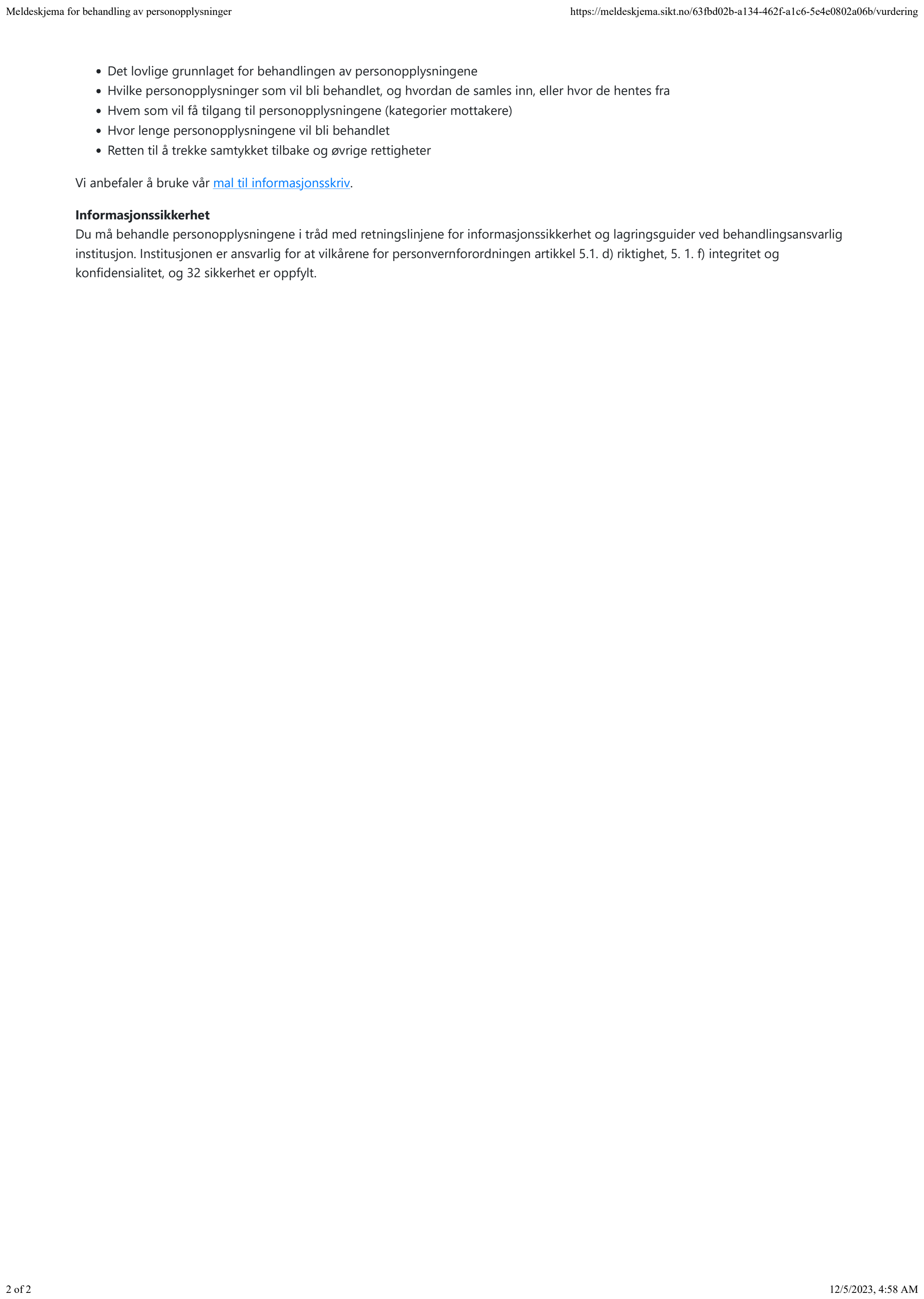
Meldeskjema

Grunnlag for automatisk vurdering
Meldeskjemaet har fått en automatisk vurdering. Det vil si at vurderingen er foretatt maskinelt, basert på informasjonen som er fylt inn i meldeskjemaet. Kun behandling av personopplysninger med lav personvernulempe og risiko får automatisk vurdering. Sentrale kriterier er:

- De registrerte er over 15 år
- Behandlingen omfatter ikke særlige kategorier personopplysninger;
  - Rasemessig eller etnisk opprinnelse
  - Politisk, religiøs eller filosofisk overbevisning
  - Fagforeningsmedlemskap
  - Genetiske data
  - Biometriske data for å entydig identifisere et individ
  - Helseopplysninger
  - Seksuelle forhold eller seksuell orientering
- Behandlingen omfatter ikke opplysninger om straffedommer og lovovertredelser
- Personopplysningene skal ikke behandles utenfor EU/EØS-området, og ingen som befinner seg utenfor EU/EØS skal ha tilgang til personopplysningene
- De registrerte mottar informasjon på forhånd om behandlingen av personopplysningene.

Informasjon til de registrerte (utvalgene) om behandlingen må inneholde

- Den behandlingsansvarliges identitet og kontaktopplysninger
- Kontaktopplysninger til personvernombudet (hvis relevant)
- Formålet med behandlingen av personopplysningene
- Det vitenskapelige formålet (formålet med studien)

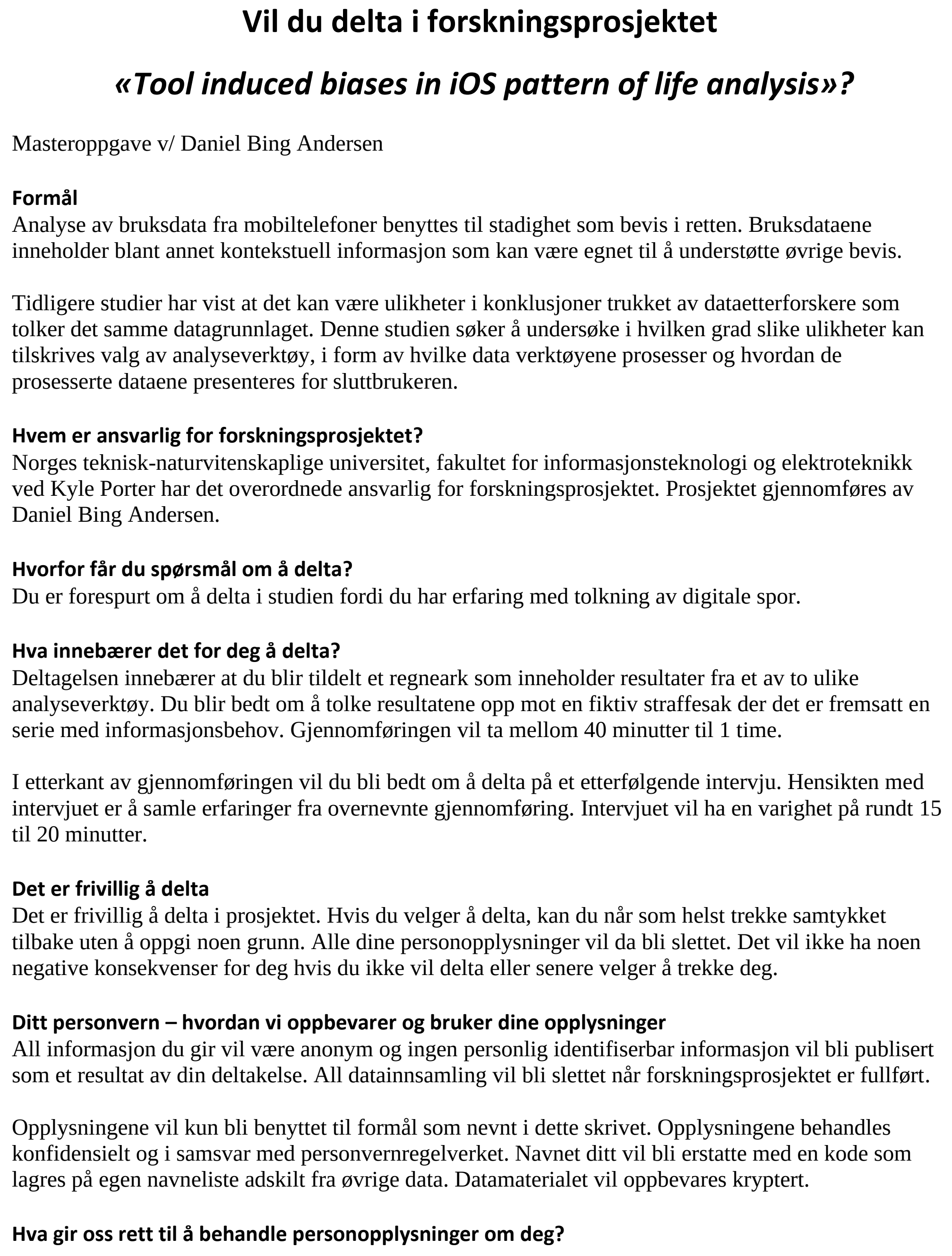


- Det lovlige grunnlaget for behandlingen av personopplysningene
- Hvilke personopplysninger som vil bli behandlet, og hvordan de samles inn, eller hvor de hentes fra
- Hvem som vil få tilgang til personopplysningene (kategorier mottakere)
- Hvor lenge personopplysningene vil bli behandlet
- Retten til å trekke samtykket tilbake og øvrige rettigheter

Vi anbefaler å bruke vår mal til informasjonsskriv.

Informasjonssikkerhet
Du må behandle personopplysningene i tråd med retningslinjene for informasjonssikkerhet og lagringsguider ved behandlingsansvarlig institusjon. Institusjonen er ansvarlig for at vilkårene for personvernforordningen artikkel 5.1. d) riktighet, 5. 1. f) integritet og konfidensialitet, og 32 sikkerhet er oppfylt.

2 Consent form



Vil du delta i forskningsprosjektet

«Tool induced biases in iOS pattern of life analysis»?

Masteroppgave v/ Daniel Bing Andersen

Formål

Analyse av bruksdata fra mobiltelefoner benyttes til stadighet som bevis i retten. Bruksdataene inneholder blant annet kontekstuell informasjon som kan være egnet til å understøtte øvrige bevis.

Tidligere studier har vist at det kan være ulikheter i konklusjoner trukket av dataetterforskere som tolker det samme datagrunnlaget. Denne studien søker å undersøke i hvilken grad slike ulikheter kan tilskrives valg av analyseverktøy, i form av hvilke data verktøyene prosesser og hvordan de prosesserte dataene presenteres for sluttbrukeren.

Hvem er ansvarlig for forskningsprosjektet?

Norges teknisk-naturvitenskaplige universitet, fakultet for informasjonsteknologi og elektroteknikk ved Kyle Porter har det overordnede ansvarlig for forskningsprosjektet. Prosjektet gjennomføres av Daniel Bing Andersen.

Hvorfor får du spørsmål om å delta?

Du er forespurt om å delta i studien fordi du har erfaring med tolkning av digitale spor.

Hva innebærer det for deg å delta?

Deltagelsen innebærer at du blir tildelt et regneark som inneholder resultater fra et av to ulike analyseverktøy. Du blir bedt om å tolke resultatene opp mot en fiktiv straffesak der det er fremsatt en serie med informasjonsbehov. Gjennomføringen vil ta mellom 40 minutter til 1 time.

I etterkant av gjennomføringen vil du bli bedt om å delta på et etterfølgende intervju. Hensikten med intervjuet er å samle erfaringer fra overnevnte gjennomføring. Intervjuet vil ha en varighet på rundt 15 til 20 minutter.

Det er frivillig å delta

Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykket tilbake uten å oppgi noen grunn. Alle dine personopplysninger vil da bli slettet. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger

All informasjon du gir vil være anonym og ingen personlig identifiserbar informasjon vil bli publisert som et resultat av din deltakelse. All datainnsamling vil bli slettet når forskningsprosjektet er fullført.

Opplysningene vil kun bli benyttet til formål som nevnt i dette skrivet. Opplysningene behandles konfidensielt og i samsvar med personvernregelverket. Navnet ditt vil bli erstatte med en kode som lagres på egen navneliste adskilt fra øvrige data. Datamaterialet vil oppbevares kryptert.

Hva gir oss rett til å behandle personopplysninger om deg?

Behandlingene av dataene baserer seg på ditt samtykke. På oppdrag fra Norges teknisk-naturvitenskaplige universitet (NTNU), har Sikt – Kunnskapssektorens tjenesteleverandør vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Dine rettigheter
Så lenge du kan identifiseres i datamaterialet, har du rett til:

- innsyn i hvilke opplysninger vi behandler om deg, og å få utlevert en kopi av opplysningene
- å få rettet opplysninger om deg som er feil eller misvisende
- å få slettet personopplysninger om deg
- å sende klage til Datatilsynet om behandlingen av dine personopplysninger

Hvis du har spørsmål til studien, eller ønsker å vite mer om eller benytte deg av dine rettigheter, ta kontakt med Daniel Bing Andersen på tlf. XXXXXXXX. Min veileder ved NTNU er Kyle Porter, tlf. XXXXXXXX. Personvernombud ved NTNU er Thomas Helgesen thomas.helgesen@ntnu.no XXXXXXXX.

Hvis du har spørsmål knyttet til vurderingen som er gjort av personverntjenestene fra Sikt, kan du ta kontakt via:

- Epost: personverntjenester@sikt.no eller telefon: 73 98 40 40.

Med vennlig hilsen

Kyle Porter *Daniel Bing Andersen*

Samtykkeerklæring

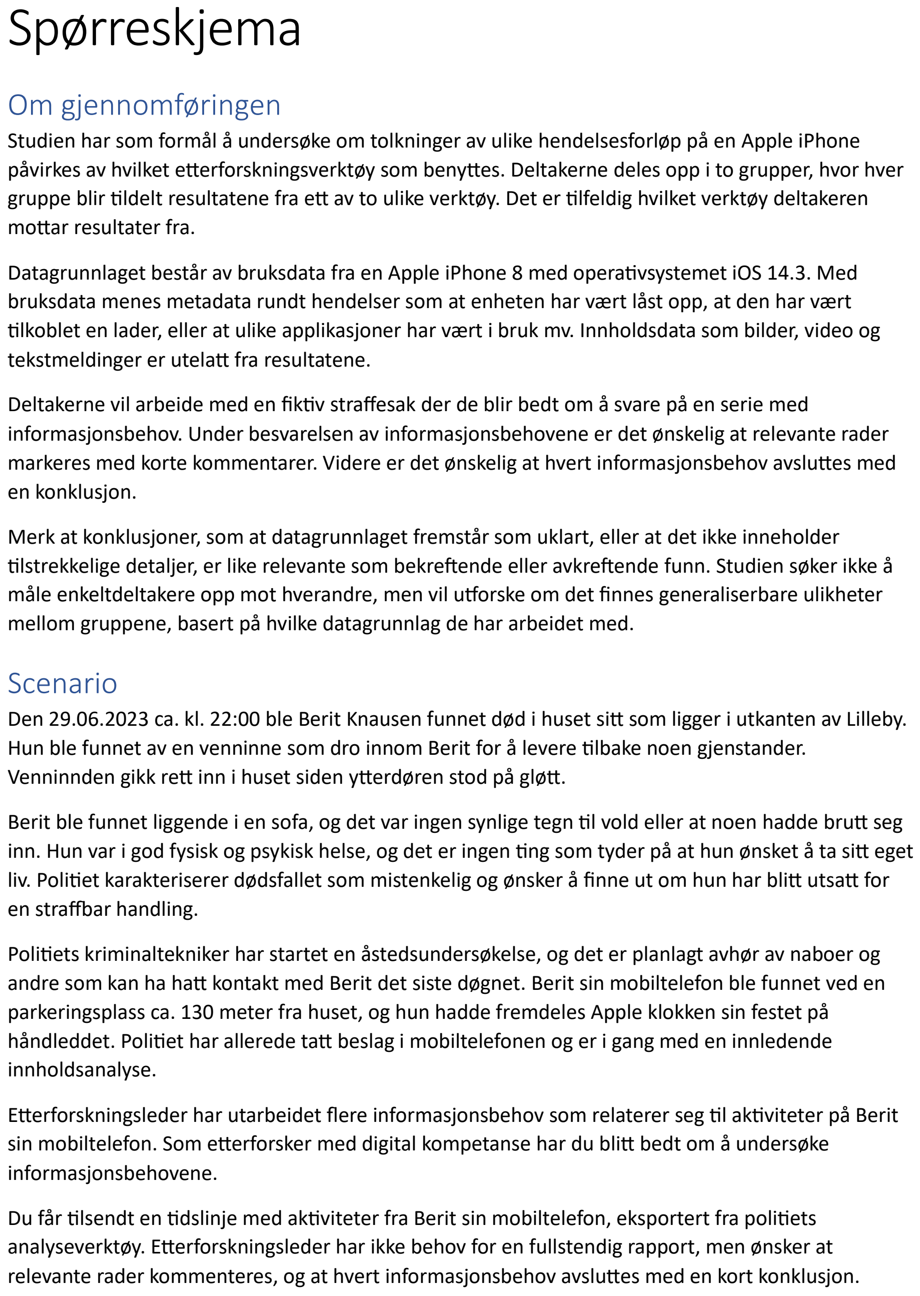
Jeg har mottatt og forstått informasjon om prosjektet, og har fått anledning til å stille spørsmål. Jeg samtykker til:

☐ å delta i saksstudien
☐ å delta i det etterfølgende intervjuet

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet

----
(Signert av prosjektdeltaker, dato)

3 Question form



Spørreskjema

Om gjennomføringen

Studien har som formål å undersøke om tolkninger av ulike hendelsesforløp på en Apple iPhone påvirkes av hvilket etterforskningsverktøy som benyttes. Deltakerne deles opp i to grupper, hvor hver gruppe blir tildelt resultatene fra ett av to ulike verktøy. Det er tilfeldig hvilket verktøy deltakeren mottar resultater fra.

Datagrunnlaget består av bruksdata fra en Apple iPhone 8 med operativsystemet iOS 14.3. Med bruksdata menes metadata rundt hendelser som at enheten har vært låst opp, at den har vært tilkoblet en lader, eller at ulike applikasjoner har vært i bruk mv. Innholdsdata som bilder, video og tekstmeldinger er utelatt fra resultatene.

Deltakerne vil arbeide med en fiktiv straffesak der de blir bedt om å svare på en serie med informasjonsbehov. Under besvarelsen av informasjonsbehovene er det ønskelig at relevante rader markeres med korte kommentarer. Videre er det ønskelig at hvert informasjonsbehov avsluttes med en konklusjon.

Merk at konklusjoner, som at datagrunnlaget fremstår som uklart, eller at det ikke inneholder tilstrekkelige detaljer, er like relevante som bekreftende eller avkreftende funn. Studien søker ikke å måle enkeltdeltakere opp mot hverandre, men vil utforske om det finnes generaliserbare ulikheter mellom gruppene, basert på hvilke datagrunnlag de har arbeidet med.

Scenario

Den 29.06.2023 ca. kl. 22:00 ble Berit Knausen funnet død i huset sitt som ligger i utkanten av Lilleby. Hun ble funnet av en venninne som dro innom Berit for å levere tilbake noen gjenstander. Venninnden gikk rett inn i huset siden ytterdøren stod på gløtt.

Berit ble funnet liggende i en sofa, og det var ingen synlige tegn til vold eller at noen hadde brutt seg inn. Hun var i god fysisk og psykisk helse, og det er ingen ting som tyder på at hun ønsket å ta sitt eget liv. Politiet karakteriserer dødsfallet som mistenkelig og ønsker å finne ut om hun har blitt utsatt for en straffbar handling.

Politiets kriminaltekniker har startet en åstedsundersøkelse, og det er planlagt avhør av naboer og andre som kan ha hatt kontakt med Berit det siste døgnet. Berit sin mobiltelefon ble funnet ved en parkeringsplass ca. 130 meter fra huset, og hun hadde fremdeles Apple klokken sin festet på håndleddet. Politiet har allerede tatt beslag i mobiltelefonen og er i gang med en innledende innholdsanalyse.

Etterforskningsleder har utarbeidet flere informasjonsbehov som relaterer seg til aktiviteter på Berit sin mobiltelefon. Som etterforsker med digital kompetanse har du blitt bedt om å undersøke informasjonsbehovene.

Du får tilsendt en tidslinje med aktiviteter fra Berit sin mobiltelefon, eksportert fra politiets analyseverktøy. Etterforskningsleder har ikke behov for en fullstendig rapport, men ønsker at relevante rader kommenteres, og at hvert informasjonsbehov avsluttes med en kort konklusjon.

Informasjonsbehov

Oppgave 1: Analyse av kommunikasjon

Etterforskningen er i sin innledende fase og det er usikkert om Berit kommuniserte med noen dagen hun døde. Du blir derfor bedt om å finne relevant kommunikasjon i tiden før hendelsen.

Oppgave 2: Tilkobling til nettverk

Det ble funnet en trådløs ruter i Berit sin leilighet. Etterforskningsleder ber deg om å undersøke om mobiltelefonen har vært tilkoblet noen trådløse nettverk og i hvilke tidsrom tilkoblingene fant sted. Håpet er at dette kan gi informasjon om når enheten sist ankom og forlot leiligheten. Det er ønskelig med ytterligere informasjon om nettverkenes SSID og BSSID.

Oppgave 3: Tilkobling til Blåtann

Etterforskningslederen ønsker at du undersøker datagrunnlaget med tanke om mobiltelefonen var tilkoblet noen Blåtann-enheter. Dersom det blir funnet noen slike tilkoblinger, ønskes det detaljer rundt hva slags enhet dette var og hva den ble benyttet til.

Oppgave 4: Kjøretøy

En av Berit sine venninner har i telefonavhør forklart at hun så Berit berit sitte i en bil samme dag som hun ble funnet død. Etterforskningsleder ønsker at du undersøker om detaljer på telefonen kan verifisere eller falsifisere denne påstanden.

Oppgave 5: Netthistorikk

Du blir bedt om å undersøke netthistorikken på Berit sin mobiltelefon. Finnes det aktiviteter som kan si noe om Berit sine handlinger i tiden før hun ble funnet død?

Oppgave 6: Opprinnelse av foto 1

Under den innledende innholdsanalysen ble det funnet et bilde som har tidsstempel kl. 18:07. Bildet er uklart, men man kan skimte det som kan være en person i bakgrunnen. Etterforskningslederen ønsker at du undersøker om bildet ble mottatt eller tatt på enheten, da dette kan gi en indikasjon på om hun var i sammen noen i tiden før hun døde.

Oppgave 7: Tilkobling til lader

Politiets kriminalteknikere har lagt merke til at ladekabelen til Berit sin mobiltelefon ligger på gulvet, ca. 1 meter fra ladeenheten som fremdeles står i stikkontakten. Kriminalteknikeren anser dette som et mulig situasjonsspor. Etterforskningsleder ber deg derfor om å undersøke tidsrommet for når enheten sist stod på lading.

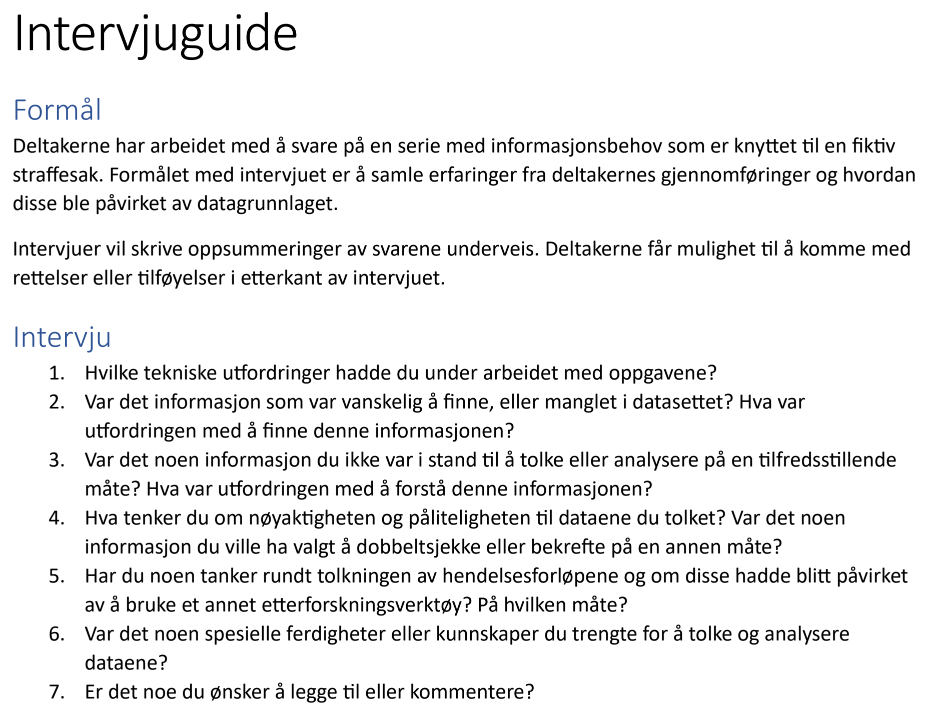
Oppgave 8: Siste aktiviteter

Det nøyaktige tidspunktet for Berit sin død er usikkert. Etterforskningledereen ber deg om å identifisere aktiviteter egnet til å indikere Berit sine siste tegn til liv. Dette inkluderer siste interaksjoner med enheten, så vel som relevante helsedata.

Oppgave 9: Flymodus

Når telefonen ble mottatt på Lilleby politistasjon var flymodus aktivert. Det er imidlertid uvisst om det var politipatruljen som tok mobiltelefonen i beslag som aktiverte flymodus, eller om telefonen allerede stod i flymodus da den ble funnet. Etterforskningsleder ønsker derfor at du undersøker når enheten sist ble satt i flymodus.

4 Interview guide



Intervjuguide

Formål

Deltakerne har arbeidet med å svare på en serie med informasjonsbehov som er knyttet til en fiktiv straffesak. Formålet med intervjuet er å samle erfaringer fra deltakernes gjennomføringer og hvordan disse ble påvirket av datagrunnlaget.

Intervjuer vil skrive oppsummeringer av svarene underveis. Deltakerne får mulighet til å komme med rettelser eller tilføyelser i etterkant av intervjuet.

Intervju

1. Hvilke tekniske utfordringer hadde du under arbeidet med oppgavene?
2. Var det informasjon som var vanskelig å finne, eller manglet i datasettet? Hva var utfordringen med å finne denne informasjonen?
3. Var det noen informasjon du ikke var i stand til å tolke eller analysere på en tilfredsstillende måte? Hva var utfordringen med å forstå denne informasjonen?
4. Hva tenker du om nøyaktigheten og påliteligheten til dataene du tolket? Var det noen informasjon du ville ha valgt å dobbeltsjekke eller bekrefte på en annen måte?
5. Har du noen tanker rundt tolkningen av hendelsesforløpene og om disse hadde blitt påvirket av å bruke et annet etterforskningsverktøy? På hvilken måte?
6. Var det noen spesielle ferdigheter eller kunnskaper du trengte for å tolke og analysere dataene?
7. Er det noe du ønsker å legge til eller kommentere?

5 Activity types APOLLO

Table 25: Activity types and rows for APOLLO

Data Type	Count
Device/App Assertions	29078
Routined Location	8200
WiFi Location	4602
App Usage Video	1783
Kernel Task Monitor	885
Coalition Interval	864
Battery Level	853
Health - Step Count	739
Motion State History	604
Process Data Usage	582
Screen Brightness	570
Video CM File	555
App Location Usage	544
Video VT Session	528
App Usage	396
Springboard Screen State	308
Network Usage	296
WiFi Connection	289
Health Heart Rate	268
Location Technology	238
Activity States	213
Cellular Location	200
Now Playing	169
Process ID	169
Device Volume	148
Activity Level	133
Widget Refresh	132
Display	124
Backlight Status	121
Health Steps	117
App Usage by Hour	117
Bluetooth State	101
Airdrop Connection	90
Data Usage	87
Application Usage	87
Battery Level UI	85
Application In Focus	67
Lock State	58
Audio Routing	57
Device Lock Imputed	57
Device Lock Status	57
Keybag Lock Status	43
Inferred Motion	42
Audio Input	33
Accessory Connection	32
DASD Activity Profile	32
App Now Playing	30
Audio Output	29
Screen Unlock State	29
Screen Time - App (By Hour)	29
Screen Time - Category (By Hour)	26
DASD Battery Temperature	24

Continues on the next page

Table 25

Data Type	**Count**
Bluetooth Connected	24
Camera State	22
Routined Location - Entry	22
Safari Browsing	16
Health Stood Up	14
Screen Time - Counted Item	13
CarPlay Connection Status	13
Screen Time - Generic (By Hour)	10
Notification Usage	10
Application Intents	10
Routined Location - Vehicle Park History	10
Health Flights Climbed	9
Telephony Registration	9
Portrait Entity	9
AWDL State	7
Push Message Received	7
Device Plugin Status	7
DASD Control Effort	7
Disk Subsystem Access	5
SMS Chat	5
Contact Interaction	5
System TLC	4
Application Web Usage	4
Widget View	3
Note History	3
Safari Activity	3
SMS Chat - Message Read	3
Knowledge Sync Addition Window	3
Routined Location - Visit Exit	3
Routined Location - Learned Location of Interest Exit	3
Application Activity	3
Knowledge Sync Deletion Bookmark	3
Location	3
Routined Location - Inbound Stop	3
Routined Location - Visit Entry	3
Routined Location - Learned Location of Interest Entry	3
Routined Location - Learned Location of Interest Transition Stop	3
Routined Location - Map Item Creation	3
Routined Location - Outbound Stop	3
Routined Location - Learned Location of Interest Transition Start	2
Routined Location - Inbound Start	2
Routined Location - Outbound Start	2
Lightning Connector Status	2
Photos	2
Siri Service	1
Application Install	1
Routined Location - Vehicle Parked	1
App Permissions	1
Airplane Mode	1

NTNU

Norwegian University of
Science and Technology