KTH ROYAL INSTITUTE
OF TECHNOLOGY

Degree Project in Technology

First cycle, 15 credits

Forensic Analysis of Footstep Data

AMIRAN DUGIEV
HENRIK CASSÉ

Stockholm, Sweden, 2023

Forensic Analysis of Footstep Data

AMIRAN DUGIEV

HENRIK CASSÉ

Degree Programme in Computer Engineering
Date: June 2, 2023

Supervisors: Leif Lindbäck, Kjell Svedman, Emil Warnicke
Examiner: Fadil Galjic

Host company: MSAB
Swedish title: Forensisk analys av fotsteg data

Abstract

Digital forensics is a niche field of study that encompasses such things as extraction, analysis and presentation of digital information, that could be used to produce forensic evidence. There are several companies whose sole specialization is providing technical software solutions to detectives to help them quickly analyze retrieved devices that might contain evidence, instead of having to send the devices to forensic labs.

However, there are still many areas that aren't fully explored within the digital forensics industry, such as using personal health data. A factor that confounds this problem is that there are many different mobile devices running different operating systems and different versions of different applications. This report examines footstep data extraction and visualization on Android. Whether this data could be used as evidence according to law enforcement agencies is also investigated.

Following a literature study to gain knowledge of the field of digital forensics, an experiment was conducted to gather data on a device through the Samsung Health application. This data was extracted and converted into a database, which was visualized using a prototype in the form of charts. Finally a technical trainer and former police officer was interviewed regarding whether the prototype could be seen as a proof-of-concept for future implementation among digital forensics solution providers. It was concluded that step data visualized in the form of graphs would be useful as forensic evidence for law enforcement detectives and juries.

Keywords

Digital forensics, Samsung Health, Footstep data, Data visualization, Android

Sammanfattning

Digital forensik är ett studieområde som omfattar extraktion, analys och presentation av digital information, som kan användas för att producera forensiskt bevis. Det finns flera firmor som specialiserar i att utrusta detektiver med mjukvarulösningar som kan analysera enheter som kan innehålla bevis, istället för att skicka enheten till ett forensiskt lab.

Det finns dock många områden som inte är helt utforskade inom digital forensik, som användning av personlig hälsodata. En faktor som utökar detta problem är att det finns många olika mobila enheter som kör olika operativsystem med olika versioner av olika applikationer. Denna rapport undersöker fotstegsdata extraktion och visualisering på Android. Om denna data kan användas som bevis av rättsväsende blir också utforskat.

Efter en litteraturstudie för att få mer kunskap inom digital forensik, utfördes ett experiment för att samla data på en enhet genom Samsung Health-applikationen. Detta data extraherades och konverterades till en databas, som visualiserades genom en prototyp i form av grafer. Slutligen intervjuades en teknisk tränare och ex-polis för att se om prototypen skulle kunna ses som ett bevis på att digital forensik skulle kunna implementera stöd för fotstegsdata i framtiden. Slutsatsen drogs att stegsdata visualiserat i form av grafer skulle vara användbara som forensiskt bevis för detektiver och juryer.

Nyckelord

Digital forensik, Samsung Health, Fotstegsdata, Data visualisering, Android

Acknowledgements

First, we would like to thank our supervisor and examiner from KTH, Leif Lindbäck and Fadil Galjic, for providing us with useful feedback during the degree project.

Second, we would also like to thank our supervisor, Emil Warnicke, and ex-supervisor, Kjell Svedman, for providing software, hardware and insights into the digital forensics industry.

Third, we would like to thank James Eichbaum for agreeing to be interviewed and for offering knowledge regarding law enforcement practices.

Stockholm, June 2023
Amiran Dugiev Henrik Cassé

Contents

List of Figures

List of Tables

List of acronyms and abbreviations

- **CSV**: Comma-seperated values
- **KTH**: Kungliga Tekniska Högskolan
- **MSAB**: Micro Systemation
- **MVVM**: Model-View-ViewModel
- **NIST**: National Institute of Standards and Technology
- **SD**: Secure Digital
- **SIM**: Subscriber Identity Module
- **SQL**: Structured Query Language
- **WPF**: Windows Presentation Foundation
- **XAML**: Extensible Application Markup Language
- **XAMN**: Examine
- **XRY**: X-Ray

Chapter 1

Introduction

Smartphones have come to dominate consumer markets and it is rare to find an individual who doesn't own one in the modern age. Smart devices generate large repositories of user data by tracking various elements about their users. Because of this they have become of great interest for law enforcement agencies due to the amount of information they can provide about potential suspects, and their ability to confirm or deny alibis ([1], page 1).

In order to meet the demand for extracted information, technical service providers have developed digital forensic tools. Digital forensic tools allow law enforcement agencies to extract and review digital evidence on-site, instead of waiting for several weeks for reports from remote digital forensic teams. Digital evidence could be things such as e-mails or online messages linking two culprits, or a web browsing history that focuses on criminal acts. They can also provide additional context based information when the use of the device is investigated, for example a criminal's device containing files on bomb-creation ([1], page 1).

A recent example of the importance of mobile forensics is french law enforcement uncovering multiple criminal rings all throughout Europe using the encrypted chat service Encrochat, by monitoring its database and giving access to the hacked chat channels to police agencies throughout the world [2].

1.1 Background

Micro Systemation (MSAB) is a company that specializes in developing mobile forensics tools and assists various institutions with digital forensics.

MSAB started as a telecommunications startup in 1984, but transitioned to focusing on digital forensics in 2003, following talks between leadership and swedish police [3].

MSAB's main software offerings are two applications, named XRY and XAMN. XRY is mainly used by on-site law enforcement officers; it can be used to extract information from smartphones. XAMN has the ability to visually represent this extracted information to find relevant evidence, such as by using filters, creating graphs or looking for specific keywords [3].

This project will be related to improving MSAB's forensic analysis concerning footstep metrics using data extracted from Android health apps, such as Samsung Health. This can make debunking criminal alibis easier and give law enforcement agencies a broader view of suspect events.

1.2 Problem

In IT-forensics, data about the individual is of great interest to produce forensic evidence. The key issue here being to produce this evidence. In this report, we will focus on investigating the possibility of producing evidence based on personal health data acquired from smartphones.

The thesis aims to answer the following research question:
How can footstep data be extracted from Android smartphones and visualized in a manner suitable for forensic analysis?
The hypothesis is that this is possible, provided sufficient technical and analytical effort.

One of the greatest ethical issues regarding this project is to which degree the suspect's privacy will be violated. Personal health data is, as the name suggests, personal. Extracting and using this data in cases involving a person might cause a violation of privacy, if not handled with care. Since the field of IT-forensics is subject to law, this is also naturally something the thesis project will have to adapt to.

Sustainability is not considered in the report as it was deemed not to be interesting. This is because the authors do not see any clear sustainability goals aligning with the research question.

1.3 Purpose

The purpose of this thesis is to improve the process of producing incriminating evidence based on personal health data, specifically footstep metrics, acquired from smartphones in a forensic examination. If successful this would be of great value for forensic examiners, in the way that their tools become more efficient to use, and for juridical personnel since personal health data can be invaluable evidence in court.

1.4 Research methodology

Initially a literature study will be done into the field of smartphone IT-forensics to gather sufficient knowledge to be able to continue the project. Also studies of the forensic software provided by MSAB will be carried using the internal resources of MSAB.
Footstep data gathering will be carried out by developing a C# prototype for extraction of data and analysis of footstep metrics. The evaluation of this prototype's quality in regards of technical viability and usability will be performed by interviewing representatives at MSAB with law enforcement experience.

1.5 Delimitations

The data analyzed in the thesis project will be limited to footstep metrics. Other type of data such as chat history, download history or activity on the phone (browsing the internet, typing in notebooks) will not be included.

The prototype used in the project will only be available in the C# language. Any software that might be needed in forensic activities (such as extracting data from ”mock phones”) will be provided by MSAB. It is also worth mentioning that the phones provided by MSAB run Android, and so the prototype for analyzing the data will be constructed for Android.

1.6 Structure of the thesis

The thesis is structured as such:

- Chapter 2 presents relevant background information about IT-forensics, specifically in the context of smartphones. This chapter will also briefly cover the forensic software given to us by MSAB. It will also give the reader a better understanding of what personal health data is and present findings from earlier studies.
- Chapter 3 presents the research methodology and each step in the research process used to solve the problem.
- Chapter 4 contains information about how data was generated and visualized. Footstep generating experiments, mobile data extraction, creation of a visualisation prototype and an expert interview are explained.
- Chapter 5 will present the results of the theses project as well as any analysis of these results. Results from the experiments are showcased, charts generated by the prototyped are presented and the answers from the interview are paraphrased.
- Chapter 6 attempts to answer the research question using the information acquired. It also discusses the work done and the project as a whole, as well as potential future application.

Chapter 2

Background

This chapter begins with giving background information about digital forensics, including smartphone forensics as well as application data in Android, legal ramifications as well as ethical and social issues. The chapter then introduces the main digital forensic tools used in the thesis. Next it is defined what ”personal health data” is and how it can be useful in digital forensics. The chapter concludes with an overview of related work.

2.1 Digital forensics

Digital forensics is ”the examination of digital storage and digital environments in order to determine what has happened” ([4], page 3). In other words, digital forensics is an investigation for digital traces of events. It should, however, be noted that digital forensics is not limited to just determining past events. It could also be used to determine current events ([4], page 3).

In digital forensics, the interesting findings are usually referred to as ”artifacts” (digital evidence) ([4], page 24). These ”artifacts” can be anything from pictures or files to chat-logs or GPS coordinates.

2.1.1 Mobile forensics

Mobile forensics is a branch of digital forensics, focusing on extracting and analyzing data obtained from mobile devices [5]. Data that can be extracted from a mobile device can be divided up into *Messaging Data*, *Device Data*, *SIM Card Data*, *Usage History Data*, *Application Data*, *Sensor Data* and *User Input Data* ([1], pages 250-251). Out of these, *Application Data* is the most interesting for this thesis.

Application data is temporary or permanent data that is used during an applications execution. This includes databases, files, images and videos ([1], page 251) [5].

Android uses a customized version of the Linux file system hence the two are very similar. [5]. The majority of the application data is stored under the */data/* folder. Preinstalled applications store their data under the */data/system/app* folder and third party apps store their data under the */data/data/app* folder. Applications may also store data in the Secure Digital (SD) card of the phone.

2.1.2 Legal, ethical and social issues

According to the Cambridge Dictionary, *privacy* is ”someone’s right to keep their personal matters and relationships secret” [6]. Since forensic investigations require access to the data and the device to be investigated, there is a breach of the users privacy. Data such as images, videos or files are all potential artifacts in a forensic investigation. There is also data that is generated ”passively” in these devices that, if combined, may paint a picture of the users personal life ([7], page 44). An example of such data could be location data generated from weather apps.

There also exist legal issues within the forensic process. For example, one of the techniques used in forensics of Android phones is called ”rooting” [5]. This technique is used to gain administrator privileges on an Android phone, in other words becoming the ”root” user. Since rooting the phone gives privileges that are normally outside of what normal users have access to, any insurance or warranty on the phone will become null.
Another legal issue regards the artifacts acquired during a forensic investigation. If these artifacts are presented in court as evidence, the artifacts have to satisfy some requirements [5]. The artifacts have to be:

- *Admissible*: Evidence must be gathered and preserved in such a way that it can be presented as evidence in court
- *Authentic*: The evidence must be relevant to the case and have a clear and relevant origin
- *Complete*: Evidence presented must be clear and complete in a way so that it shows the whole story
- *Reliable*: Techniques used and evidence collected must not infringe on the authenticity of the evidence
- *Believable*: The evidence presented must be clear, easy to understand and believable

If these requirements are not met, the artifacts may be ruled as inadmissible. This is a big issue in mobile forensics. Since data can be accessed, synchronized and stored across multiple devices it is difficult to keep track of origin. Another issue arises from the fact that the data can be accessed remotely by an adversary which may sabotage the evidence. A common trick used by forensic experts to combat this is to turn on the ”airplane” mode on the phone, which turns off networking.

2.2 Digital forensic tools

In order to generate footstep metrics this examination will be using two digital forensic tools. XRY to extract data from a test device and XAMN to examine and filter this information.

2.2.1 XRY (X-Ray)

XRY is a digital forensic tool developed by MSAB, whose primary purpose is extraction of data from mobile devices. The application is designed for use by trained forensic specialists to be able to recover relevant digital content fast and reliably, while retaining its integrity. XRY has two main versions, called LOGICAL and PHYSICAL [8].

XRY LOGICAL allows a user to access and recover file system data on the crime scene by communicating with the operating system of the mobile device. It also reads the Sim and SD cards for additional hardware related information [8].

XRY PHYSICAL bypasses interfacing with the operating system entirely and instead extracts raw data from the device. This memory dump contains system, protected and deleted data. Having access to raw data also lets you bypass certain security restrictions that operating systems set [8].

XRY has some other more specific software functions. XRY PINPOINT can detect pin configuration for non-standard devices, such as cheap imitation phones. XRY CAMERA is a hardware solution that allows you to take images and video of the device, which can for example help with identifying oil and dirt marks. XRY CLOUD can extract information on the cloud by extracting credential tokens on the device, which allow the device to verify its connections to cloud based storage centers. Finally XRY EXPRESS offers the functionality of XRY, but through a simplified user interface to make it possible for non-experts to use [8].

2.2.2 XAMN (Examine)

XAMN is a software solution developed by MSAB for digital forensics analysis. According to MSAB a tool for analyzing devices is required due to mobile smart devices being able to contain far more data than older mobile devices. The data can consist of thousands of messages, images and other files, which can constitute several gigabytes of data. This large amount of data presents challenges towards being able to find data that might be relevant as evidence, without presenting a massive opportunity cost to investigators. For this reason tools such as XAMN have been developed to allow efficient searching, filtering and analysis of data [9].

To start searching for data using XAMN, investigators first upload data that was retrieved from XRY. Once this is completed relevant data can be crunched down by filtering it against such criteria as location, content category, time, phone number used, file size and more. The analytical view is a modular graphical interface that allows a user to display the filtered data in the form of galleries of images, conversations between people, relationships between people, geography or more simple lists. These views are generated due to XAMN examining links between data, such as someone sending messages to someone else and where they sent these messages to form connections between different users [9].

Some additional features of XAMN is the detail view, which can show more in-depth data about a selected element such as the geographic area or timeframe it was created in. There are also tools to simplify the filtering process with ”quick-views” which are pre-created filters that are built into the application, such as only showing recent calls or messages. The user has the option to create their own quick view. The user can tag different data as ”important” or ”unimportant” or create their own tags. Finally everything that the user has found and analyzed can be compiled together into a comprehensive ”report” file with all elements deemed relevant for a particular investigation [9].

2.2.3 Forensic tool reviews

XAMN and XRY have both been tested by the National Institute of Standards and Technology (NIST) [10]. NIST is a part of the U.S. Department of Commerce, with the primary mission to promote innovation and industrial competitiveness by advancing scientific and technological standards in the U.S. [11]. In the testing NIST conducted on XRY and XAMN, all test data was acquired completely and accurately for all tested mobile devices, except for some anomalies. The anomalies are:

- An error message is not reported to the user in case of connectivity disruption for some devices
- Internet related data (i.e., Bookmarks, History, Email) is not reported for some devices
- GPS data (i.e., latitute, longitude coordinates) is not reported for some devices

The anomalies listed above are the ones that concern Android devices. Anomalies for iPhone or other non-relevant devices are not listed due to brevity. It should also be mentioned that some of the anomalies reported in [10] have been excluded because MSAB have addressed and solved these anomalies.

2.3 Personal health data

This section elaborates on how personal health data is acquired and used in the medical field by means of mobile health applications. It also examines some examples of data that could be extracted from those applications in previous studies.

2.3.1 Mobile health applications

Due to advances in mobile technology, smart devices have become powerful health monitoring tools. Patients in remote areas can be monitored through the use of a mobile devices and periodic reports from mobile sensors can be sent through wireless communication to a central monitoring center, set up by a collaboration between various medical institutes ([12], page 2).

Mobile health systems save hospitals and other medical establishments time and resources, since electronic patient record updates can be automated and less personnel is required to adequately monitor patients' needs ([12], page 2).

In order to provide supply for the demand of mobile technologies that improve medical systems, mHealth has become a growing industry. mHealth is a term for the practice of medicine and public health supported by mobile technologies. As of 2017 there were more than 325 000 mHealth applications available, with Android being the leading market ([13], page 2).

mHealth does however pose a significant security and privacy risk towards patients, due to the possibility of hacking and retrieving sensitive personal data and information from such applications ([13], page 2).

2.3.2 Examples of extracted data

In one example XRY was used to scrape test data generated by the authors. The data included name, e-mail, age, gender, birthday, zip code, height, weight, lactation, ethnicity, blood type, waist measurements, neck measurements, hip measurements, shoe measurements, shoe brand used, detailed medicine usage, routes walked, DNA records, and blood pressure. This type of data is extremely attractive to law enforcement agencies, since it exposes the physical appearance of the culprits on top of revealing their identity. It is also possible to find passwords which would allow hackers to hack alternative avenues such as social media or online governmental services since a majority of online users re-use passwords. Some of the information was even obtainable through simple leaderboard web scraping ([13], page 4).

In a second example XRY was used on the health application MyFitnessPal to extract passwords in plain text and the 4 digit pin to unlock the smart device itself as well as various exercise and diet information. Data from the RunKeeper app was extracted with enough information to map out a complete map of each of the user's trips, including deleted entries. Lastly data from Period Calendar was extracted, which contained information about contraceptive use and menstrual cycle timings ([14], page 3).

2.4 Related work

This section contains previous related work that is interesting for the thesis.

2.4.1 Investigating footstep data as forensic evidence

In the paper [15], Jan Peter van Zandwijk and Abdul Boztas investigate footstep data, extracted from the iPhone's inbuilt health application *Health App*, as a potential source for digital evidence. The three versions of the iPhone used in the paper were the iPhone 6 (iOS 10.2 and 10.3.3), iPhone 7 (iOS 10.3.1) and iPhone 8 (iOS 11.1.2).

Method used in the paper

Five subjects were asked to travel certain distances by walking along a defined route. This route was divided up into three parts of different length. The subjects traveled these three parts at least two times in walking pace and at least two times in a "freely chosen moderate running pace", while carrying iPhones in different parts of their clothing (trouser and jacket pocket), a backpack and their hands. These tests were done both with and without specifying biometric information (such as weight and height) in the *Health App* user interface. For each test, the amount of steps taken were manually recorded using a tally counter by both the subject and the experimenter. Start and end time of the test was also recorded using a calibrated watch. After each measurement session the registered data from *Health App*, for each test, was extracted using manual SQL commands. The extracted times were summed up.

Major results presented in the paper

The deviation between the measured number of taken steps and the amount of steps registered by *Health App* in all of the iPhones was small and the measurement abilities of *Health App* was deemed accurate.

The registered distance traveled seemed to depend on where the iPhone was carried, the walking speed, as well as the subjects. This was assumed by the authors to be because of variations in the way the subjects walked. Data extracted from the iPhone 7 indicates great variation between the registered distance and the measured one, and in the majority of the cases the registered distance is lower. Significant variations in distance traveled could also be found between subjects. The reason for these variations, the authors argued, was due to the amount of forward-backward movement subjects performed during locomotion. With biometric data specified in *Health App*, a slight increase in registered distance, compared to without it, was observed.

Minor results presented in the paper

A variation in time intervals in which the data is stored in *Health App* was noticed between the different versions of the iPhones. For the iPhone 6 the most common time interval was 70 seconds and for the iPhone 7 the most common time interval was 60 seconds. This implies, according to the authors, that if a trial takes more time than the time interval of data storage, the data for that trial will be stored in multiple entries in *Health App's* database. For the iPhone 8, no definite time interval for data storage was found. Additionally, the time frames for the iPhone 6 and 7 seemed, according to the authors, to be adapted dynamically. For longer activity windows, the time frame can increase up to 600 seconds.

Conclusions and usage in the thesis

One of the major results of the paper is the fact that the registered distance traveled seems to be heavily influenced by walking style, speed and carrying location of the iPhones. Using this data as forensic evidence may prove difficult due to it's uncertainty. The number of registered steps however seems to be accurate compared to the measured amount.

This report will serve as a foundational building block in the method of our work. This report shows a detailed way of accurately establishing whether distance and number of footsteps registered by health apps are accurate. The above paper investigates this accuracy on the iPhone (iOS), which means that expansion of the investigation to also cover Android may be relevant.

2.4.2 Forensic Taxonomy of Android Health Apps

In the paper [14], the authors proposed a taxonomy for forensic artifacts extracted from Android health apps. The taxonomy categorized the health apps and the artifacts. The health apps were categorized based on primary use and functionality. For example, apps that were more catered to fitness and training were categorised as *Health apps for the lay person*, while apps that were catered more for health and well-being were categorised under *Patient care and monitoring*.

The use of this paper for this report is mainly found in the detailed information given in the method section about the database structure of these health apps. This detailed structure will be useful in understanding more where and how Android health apps store their data. It should also be mentioned that the forensic tool used in the paper happened to be XRY, the same extraction tool used as in this thesis.

Chapter 3

Method

The purpose of this chapter is to provide an overview of the research method used in this thesis. The first section briefly introduces the research process. Subsequent sections describes the method involved with each step of the research process.

3.1 Research Process

The research process in the thesis consists of five steps:

- **Problem statement**: A problem statement is specified in the form of a research question.
- **Literature study**: Information relevant for the thesis is collected and presented.
- **Generating footstep data**: During the case study test data is collected in the form of footstep data extracted from popular Android health apps in experiments.
- **Visualizing footstep data**: A prototype is developed for visualizing data with the purpose of aiding in analysing the data collected in the case study.
- **Interview**: An ex-police officer is interviewed to evaluate the prototype.
- **Evaluation**: The work done is evaluated by using both quantitative and qualitative methods.

Figure 3.1 shows the process visually.

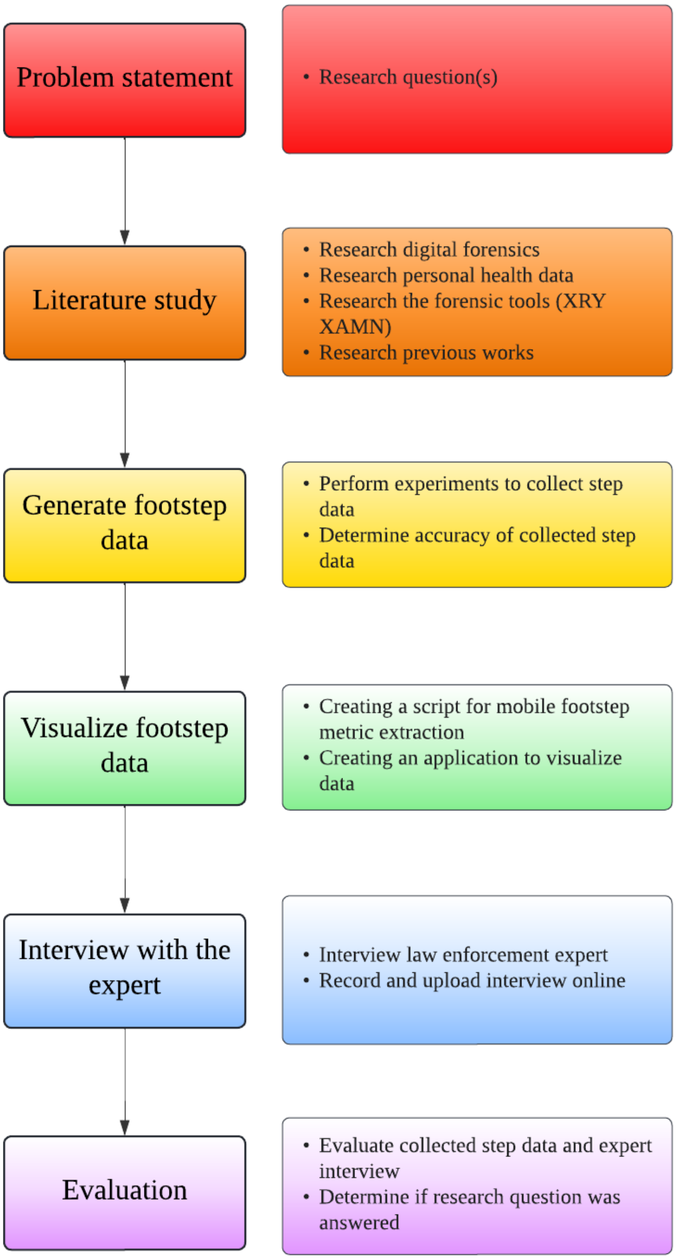

Problem statement
• Research question(s)
Literature study
• Research digital forensics
• Research personal health data
• Research the forensic tools (XRY XAMN)
• Research previous works
Generate footstep data
• Perform experiments to collect step data
• Determine accuracy of collected step data
Visualize footstep data
• Creating a script for mobile footstep metric extraction
• Creating an application to visualize data
Interview with the expert
• Interview law enforcement expert
• Record and upload interview online
Evaluation
• Evaluate collected step data and expert interview
• Determine if research question was answered

Figure 3.1: Research Process

3.2 Problem statement

The first step of the research process is to create a problem statement. To create a problem statement for the thesis, a research question is identified and specified. This research question is iteratively refined by supervision from both the supervisors at KTH and MSAB. The purpose of this step is to gain a better understanding of the research problem as well as a clearer view of the purpose of the thesis.

3.3 Literature study

The second step in the research process is a literature study. The purpose of the literature study is to gain a deeper understanding of the topics researched in the thesis [16]. Another more practical purpose of this step is to summarize the most relevant knowledge to the readers of the thesis report.

The main recommended method for conducting a literature study and locating relevant sources in the field of software engineering is querying online databases using keywords [17]. This method is adopted and the two main databases used are Google Scholar and the online publisher O'Reilly.

An effort was made to use scientific articles, but it was found that books introduced and explained subjects more concisely and therefore became the preferred sources when searching for knowledge about subjects.

For acquiring knowledge about the forensic tools, the official documentation issued by the tools' manufacturer is used. As a last resort if sufficient sources could not be found, Google is used to find information. When this approach is adopted, the sources are evaluated based on the following criteria [18]:

- *Purpose*: What is the author trying to accomplish?
- *Scope*: How much of the topic is covered and in what depth?
- *Authority*: What is the authority of the source – are the author and his or her credentials given?
- *Audience*: Who is the publication intended for?
- *Format*: How is the information presented?

3.4 Generating footstep data

The next step of the research process consists of experiments. The purpose of these experiments are to generate test data in the form of footstep data. This data will later be analyzed in the evaluation step. The experiments are carried out by one of the authors. Therefore, any ethical or social issues regarding participation will be eliminated.

The data is generated by walking through different types of terrain. The method of walking to generate footstep data is the one most used in similar works (such as Peter van Zandwijk and Abdul Boztas [15]) and is also the most intuitive. The terrain being investigated are roads and forests, as these are the two most common terrains found. The number of steps taken is manually recorded using a physical pedometer as a second reference to be compared to the prototype.

Each experiment is conducted five times for each method of data generation (five times for both terrains). Each experiment takes approximately fifteen minutes to complete. Dividing up the experiment into five smaller experiments for each terrain allow for a larger sample size to analyze and draw conclusions from. The average adult walks between 5000 to 7000 steps per day [19] [20]. This approximately corresponds to a distance of 5 kilometer (when walked by the authors). Therefore, to ensure consistency, a predefined route with a distance of 1 kilometer is walked for each execution of the experiment.

The exact data being measured depends on which application is being examined but it is hypothesized that common data that can be acquired consists of footstep amount, distance and direction.

Samsung Health has over 1 billion downloads on Google Play [21]. Moreover, Samsung Health is also installed on newer Android smartphones by default [22]. Due to its popularity, it was chosen as the health app to be investigated.

Additionally, a smaller experiment was done to evaluate the accuracy of both the pedometer as well as the Samsung Health application. This experiment consists of 10 iterations where the participant walks 50 steps straight on a flat surface, manually counted by the participant, and records the amount of steps walked shown by both the Samsung Health app as well as the pedometer.

3.5 Visualizing footstep data

In order to analyze extracted footstep metrics, a prototype is developed that can visualize the data in the form of graphs. Some samples are also manually measured to ensure there are no technical errors.

The purpose of this step is data visualization, with the purpose of aiding in analysis of the data collected during footstep data generation. Since the software being developed is a prototype, it was deemed unnecessary to follow methods designed for large scale applications.

The main form of proving the reliability of prototyping will be through test-retest. This means that the consistency of results are proven by repeating the same test. The visualization should give similar results if given the similar input, for example by walking the same path twice in a row and using both tries as test data [23].

Prototyping allows technical validation by verifying that a software solution can be used to produce evidence. Prototypes are also quick to produce, which is important, since otherwise it would take too much time away from answering the research question [24].

The prototype is developed in Visual Studio. Visual Studio is an integrated development environment developed by Microsoft. Visual Studio allows its user to write and debug code as well as many other features. The prototype was developed using Windows Presentation Foundation (WPF), which is a User Interface Framework offered by Visual Studio that allows for the creation of programmable graphical applications. C# is the programming language used for programming WPF applications. Extensible Application Markup Language (XAML) is used to structure and define User Interface elements [25].

The prototype will be making use of the WPFChart addon imported through NuGet which is a package manager built into Visual Studio [26], WPFChart contains several methods and components to enable creating charts in the application [27]. The structure of the program will follow the Microsoft recommended MVVM design pattern which consists of a view that presents objects to the user, a viewmodel that populates the view with data from the model, and the model which contains all programming methods [28].

3.6 Interview with the expert

An expert interview is performed with an MSAB trainer who was formerly a police officer in the USA. Expert interviews are seen as reliable due to there being acceptable levels of inter-expert agreement [29].

Interviews are a common qualitative data technique intended to find out the understandings, opinions and experiences of participants. For the purposes of finding an answer to the research question a structured interview will be conducted with pre-written questions, somewhat like a verbal survey. Structured interviews increase the reliability and credibility of the interview and allow for direct and concise answers to the research question [30].

The interview is conducted through Zoom. Zoom is a video communications application that allows several people to be in a group call. Zoom is used due to the person being interviewed traveling a lot between Sweden and the USA and therefore a physical interview being difficult to schedule. Zoom also has built-in video recording features that make saving the interview easier and means the interview questions and answers don't have to be stored through a potentially unreliable format such as using pen and paper. During the Covid pandemic Zoom was widely used by government, education and business institutions and is therefore recognizable and usable by most people [31].

A demo of the prototype will be shown to the expert through a streamed computer screen and prototype capabilities will be showcased, such as graphing out various footstep metrics. Afterwards there will be questioning to gauge whether the expert considers that the prototype could be useful in his line of work and if he could think of potential improvements. This interview format is chosen to acquire direct answers to the research question from someone with authority in the field of study.

The interview will be recorded and interpreted. Answers given by the law enforcement expert and what it could mean for the research question, for example whether the prototype or paper would be of interest in a law enforcement environment.

3.7 Evaluation

The next step in the research process is the evaluation of the collected data and the prototype. It is assumed that the research question may be answered in this way.

In order to analyze data obtained from the footstep data generation and the interview with the expert, a quantitative approach will be applied towards analyzing the footstep data generation, while a qualitative approach will be used concerning the interview.

This is because quantitative approaches are more suitable for analyzing data concerning elements you can measure or count, such as footstep metrics. Qualitative approaches are more applicable towards data that is comprised of signifiers that the author interprets [32].

To be able to answer the research question the footstep metrics gathered will be analyzed through the prototype. The usefulness of the developed prototype is evaluated using an interview.

The primary evaluation method is careful interpretation of the expert interview to answer the research question. The research question will be considered to be properly investigated once the expert, after getting exposed to our research through access to this report and getting a demo of our prototype, indicates that Android footstep metrics are sufficient as evidence for law enforcement or not.

Chapter 4

Acquiring necessary data

The purpose of this chapter is to provide an overview of the process involved in generating the footstep data used in this thesis. The chapter begins with detailing the experiments as well as giving an overview of the tools, such as the software and hardware used in the experiments. Next, the chapter explains the process of extracting the footstep data from the mobile test device. The database that was created using the footstep data is detailed. The structure of the prototype that visualizes the contents of that database is also described. Finally, the chapter briefly gives an overview of how the interview related to step data is conducted.

4.1 Generating data

The footstep data generation is conducted in two experiments. The first experiment simulates a walking distance of 1 kilometer on a road or flat surface. The second experiment simulates a walking distance of 1 kilometer in forest or uneven surface. Both experiments are conducted in a park.

For the experiment simulating a road, the participant walks a pre-defined route. This route consists of roads around and in the park (see figure 4.1). For the experiment simulating forest, the participant started at one end of the park and walked to the other, then walked back to the starting position again (see figure 4.2). This is due to the fact that the distance from one end to the other in the park is only around 500 meters.

These experiments are conducted five times for a total distance walked of 5 kilometers per experiment. For each execution of the experiment, the participant records the start and finish times as well as the steps walked according to the pedometer. The pedometer is placed at hip level on the left side of the body and the smart phone running the Samsung Health application is placed in the right front pocket.

The experiments are conducted one day apart, with the first experiment being conducted the 12:th of April and the second one on the 13:th. Because of this, the step counter in Samsung Health is paused between experiments, as to minimize overhead when extracting and analyzing the data later.

Three main tools are used in the experiments: one smartphone, one pedometer and the health app. The smartphone chosen for the experiments is the Samsung Galaxy S10. The pedometer used in the experiments is the "Rubicson Pedometer with belt clip". Samsung Health was chosen to be the health app to investigate.

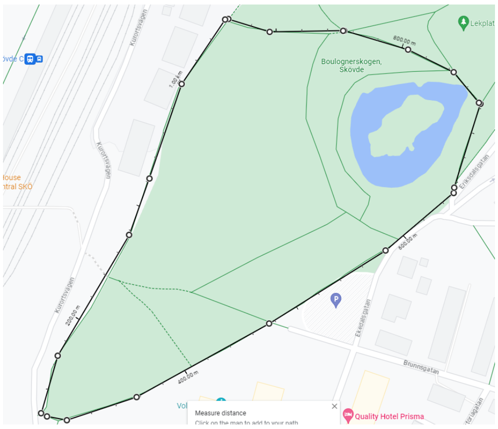


Figure 4.1: Route walked to simulate a road

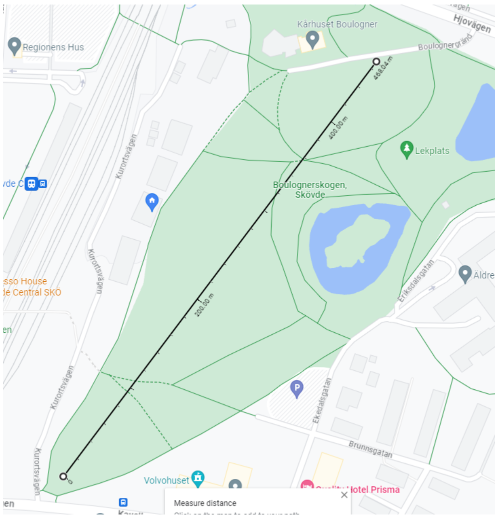


Figure 4.2: Route walked to simulate a forest

4.2 Extracting data

The extraction process began by using XRY LOGICAL on the mobile device that contained the footstep data. XRY LOGICAL attempts an extraction by examining each screen on the device and recording what is displayed. However, XRY LOGICAL does not support extraction from the Samsung Health application, so this method was found to be unsuccessful after the XRY file was imported for examination inside XAMN.

Next, XRY PHYSICAL was attempted, which bypasses the operating system of the device and dumps raw data found on the device. Since a raw dump simply extracts any data it can find instead of interacting with applications, platform support would be irrelevant. Database files were located by searching using the keyword ”SHealth” in XAMN. Filters were used in order to only find database files. However they were in an encrypted format, which couldn’t be accessed.

Finally, a manual approach was taken by downloading personal data from the settings menu inside Samsung Health. This dumped several csv files into device storage. A csv file is essentially a text file which contains values that are separated by commas, making it easy to import into and export out of databases. The csv files were examined in XAMN, and information related to footstep metrics were found.

The file ”com.samsung.shealth.tracker.pedometer_step_count.csv contained values for footsteps taken, timestamps, calories burned, speed and distance, so it was chosen as the primary source of data for visualizing. It was downloaded from the phone using XRY LOGICAL and stored in a laptop.

The csv file was then imported as a table in a SQLite file using DB Browser for SQLite, which is a program that provides a graphical interface for editing SQLite files. SQLite allows for the creation of hardware embedded databases that can be used without a database management system, it is commonly used in mobiles. The table was split up into two individual tables, one for road measurements and the other for forest measurements. The tables were then sorted using timestamps from within the application interface. Dot notations for decimals were replaced with commas to prevent formatting errors related to localization.

The aforementioned changes were executed using the query displayed in 4.3, and the table itself is displayed in 5.4:

```
UPDATE step_data_road
 set fieldX = REPLACE(fieldX, '.', ',')
```

Figure 4.3: SQL query used to order tables by time, with ”X” being a chosen field

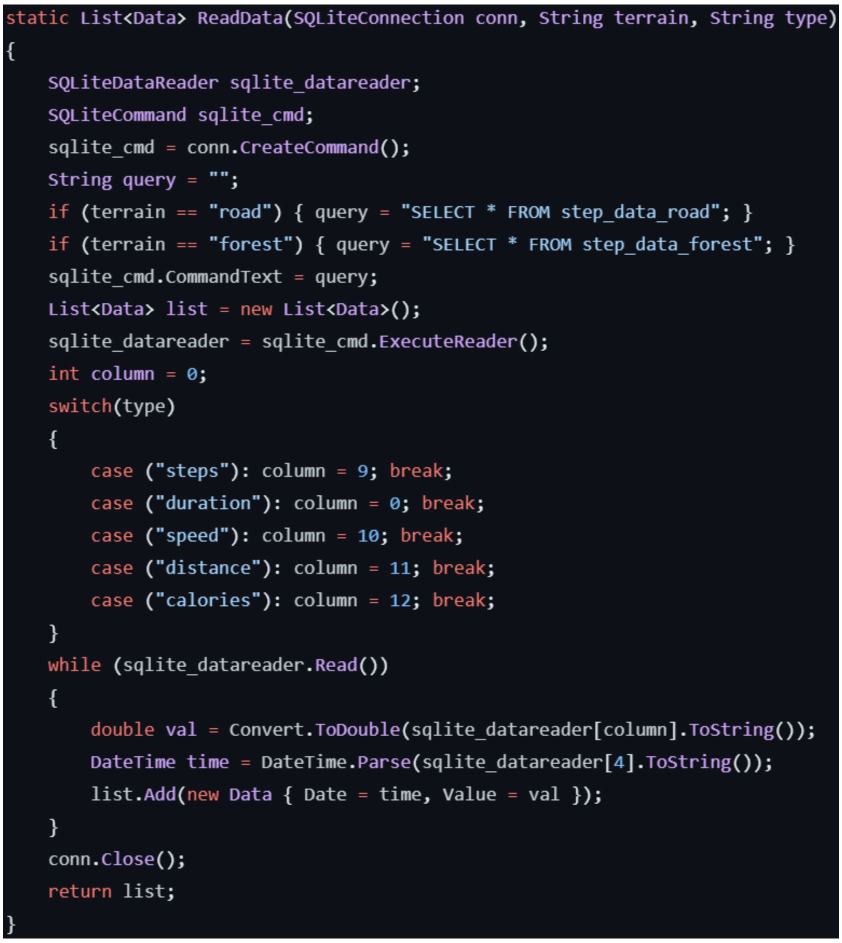
4.3 Creating a prototype

To create a prototype for the purpose of visualizing the SQLite database, a Visual Studio project was created. This project was also uploaded as a Github repository to facilitate co-operative programming and publicly display the codebase. This repository can be found at: [33].

Then, to enable the creation of charts, the WPFChart addon was imported into the project, using the NuGet package manager. WPFChart contains many methods and properties for the creation of charts in applications using a MVVM software pattern.

The main parts of the program are located in the DBInfo.cs (model) and MainWindows.XAML (view) files. ViewModel.cs (viewmodel) passes values from the model to the view. The model creates a SQLite connection during its construction. It also has parameters which allows the viewmodel to choose what kind of footstep metrics to receive, for example step data from forest metrics or calorie data from road metrics.

The model's main method is ReadData which queries the database through the SQLite connection. Depending on what method parameters are entered, a list of Data objects containing dates and values of the requested information are returned. The source code for ReadData is displayed in 4.4:



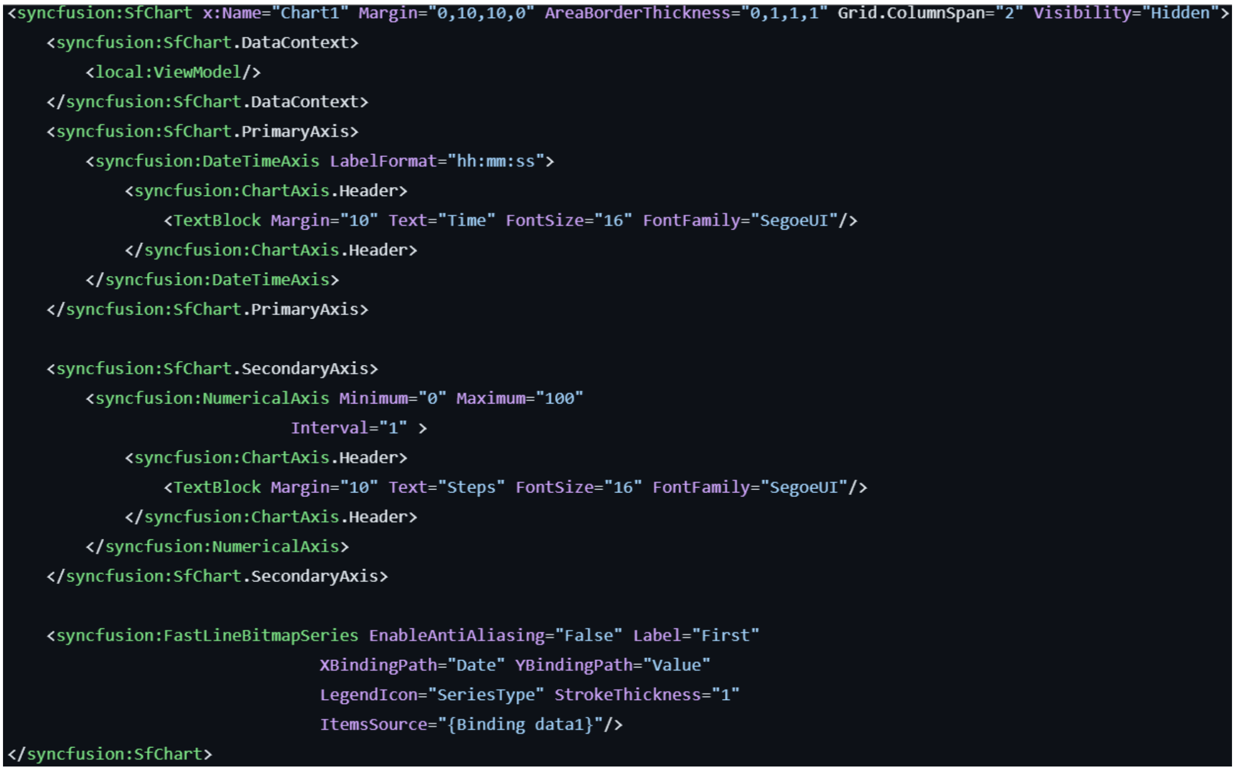
```
static List<Data> ReadData(SQLiteConnection conn, String terrain, String type)
{
    SQLiteDataReader sqlite_datareader;
    SQLiteCommand sqlite_cmd;
    sqlite_cmd = conn.CreateCommand();
    String query = "";
    if (terrain == "road") { query = "SELECT * FROM step_data_road"; }
    if (terrain == "forest") { query = "SELECT * FROM step_data_forest"; }
    sqlite_cmd.CommandText = query;
    List<Data> list = new List<Data>();
    sqlite_datareader = sqlite_cmd.ExecuteReader();
    int column = 0;
    switch(type)
    {
        case ("steps"): column = 9; break;
        case ("duration"): column = 0; break;
        case ("speed"): column = 10; break;
        case ("distance"): column = 11; break;
        case ("calories"): column = 12; break;
    }
    while (sqlite_datareader.Read())
    {
        double val = Convert.ToDouble(sqlite_datareader[column].ToString());
        DateTime time = DateTime.Parse(sqlite_datareader[4].ToString());
        list.Add(new Data { Date = time, Value = val });
    }
    conn.Close();
    return list;
}
```

Figure 4.4: The model’s ReadData method

The view uses WPFChart to create several charts by defining an X and Y axis. A line series is also defined, which together with the axes form a graph. The graph is populated by acquiring data from the viewmodel. The XAMN code required to define a chart is shown in 4.5:
There is also a selection box used to choose which chart should be displayed, the other charts being hidden from the view in the process. The selection box and generated charts can be seen in figures 5.1-5.5.

4.4 The interview

The interview has two purposes. First, to gain feedback on the prototype from an experienced authority. Second, to consult an expert on the research question



```
<syncfusion:SfChart x:Name="Chart1" Margin="0,10,10,0" AreaBorderThickness="0,1,1,1" Grid.ColumnSpan="2" Visibility="Hidden">
    <syncfusion:SfChart.DataContext>
        <local:ViewModel/>
    </syncfusion:SfChart.DataContext>
    <syncfusion:SfChart.PrimaryAxis>
        <syncfusion:DateTimeAxis LabelFormat="hh:mm:ss">
            <syncfusion:ChartAxis.Header>
                <TextBlock Margin="10" Text="Time" FontSize="16" FontFamily="SegoeUI"/>
            </syncfusion:ChartAxis.Header>
        </syncfusion:DateTimeAxis>
    </syncfusion:SfChart.PrimaryAxis>

    <syncfusion:SfChart.SecondaryAxis>
        <syncfusion:NumericalAxis Minimum="0" Maximum="100"
                             Interval="1" >
            <syncfusion:ChartAxis.Header>
                <TextBlock Margin="10" Text="Steps" FontSize="16" FontFamily="SegoeUI"/>
            </syncfusion:ChartAxis.Header>
        </syncfusion:NumericalAxis>
    </syncfusion:SfChart.SecondaryAxis>

    <syncfusion:FastLineBitmapSeries EnableAntiAliasing="False" Label="First"
                                XBindingPath="Date" YBindingPath="Value"
                                LegendIcon="SeriesType" StrokeThickness="1"
                                ItemsSource="{Binding data1}"/>
</syncfusion:SfChart>
```

Figure 4.5: WPFChart implementation using XAML

and its importance. The interview is conducted digitally via Zoom with ten predefined questions which the interviewee was informed of beforehand.

The questions are divided up among the authors, where one asks the first five questions and the other asks the last five questions. The meeting is recorded and the answers were evaluated after the interview had been concluded.

The interviewers consisted of the authors and the interviewee was James Eichbaum. James is a former U.S. police officer with a background in digital forensics and is now a training manager for MSAB. James was shown a quick demo of the prototype before being asked questions. The full video interview can be found at: [34].

Chapter 5

Results

The purpose of this chapter is to present the results that was acquired. The chapter begins with a section where the results from the experiments are shown. Next, the created charts are presented and explained. Finally, the chapter finishes by detailing the interview questions and answers acquired from the expert.

5.1 Created data

The results from the experiments are presented in the tables below. Table 5.1 and 5.2 both have four columns. The first column, *Test number*, contains the number of the test for ordering purposes. Next, *Number of steps: pedometer*, contains the steps recorded by the pedometer for each test. The third column, *Number of steps: Samsung Health*, contains the number of steps recorded by Samsung Health for each test. This data was acquired by adding each entry from the created database (see table 5.4) present within the time frame corresponding to each test. Finally, *Time frame*, contains the time frame for each test that was carried out.

Table 5.1: Results from the road experiments

Test number	Number of steps: pedometer	Number of steps: Samsung Health	Time frame
1	1207	1155	14:16 - 14:28
2	1178	1147	14:34 - 14:46
3	1222	1182	15:02 - 15:14
4	1276	1236	15:16 - 15:29
5	1373	1177	15:31 - 15:44

Table 5.2: Results from the forest experiments

Test number	Number of steps: pedometer	Number of steps: Samsung Health	Time frame
1	1450	1309	11:41 - 11:58
2	1445	1390	12:01 - 12:17
3	1284	1240	12:19 - 12:33
4	1320	1244	12:35 - 12:49
5	1340	1256	12:51 - 13:05

For table 5.3, the columns are the same except that the *Time frame* was deemed unnecessary, because each time frame was one minute or less. Another difference is that the number of steps from Samsung Health for each test was directly observed on the apps interface.

Table 5.3: Comparison between pedometer and Samsung Health for 50 steps

Test number	Number of steps: pedometer	Number of steps: Samsung Health
1	55	50
2	59	48
3	53	50
4	57	51
5	51	51
6	56	48
7	53	50
8	58	51
9	59	52
10	57	49

Table 5.4 is a representation of the first 5 rows in the SQLite table that was created from Samsung Health data. In total 157 rows were created after both the road and forest experiments. The tables contained other information, such as timezones, but they weren't considered relevant.

Table 5.4: The SQLite table created from the csv file

Entry	Steps	Duration	Speed	Distance	Calories	Time
1	7	5389	0.916	4.94	0.26	2023-04-12 12:13:00
2	57	39012	1.055	41.18	2.25	2023-04-12 12:14:00
3	89	59916	10.083	64.91	3.53	2023-04-12 12:16:00
4	92	60000	1.118	67.12	3.68	2023-04-12 12:17:00
5	92	60000	1.118	67.09	3.68	2023-04-12 12:18:00

5.2 Created charts

The prototype's graphical interface is showcased. Charts denoting each value that is visualized, such as steps (5.1), duration (5.2), speed (5.3), distance (5.4) and calories (5.5), are shown. All charts use data from road experiments, but data using forest experiments are also selectable from the drop-down combobox, meaning there are 10 chart options total. Each chart except the one for speed has 5 sharp drops due to measurements being written down after a 1km walking test. Speed has only a single sharp drop due to the step counting being paused to adjust equipment.

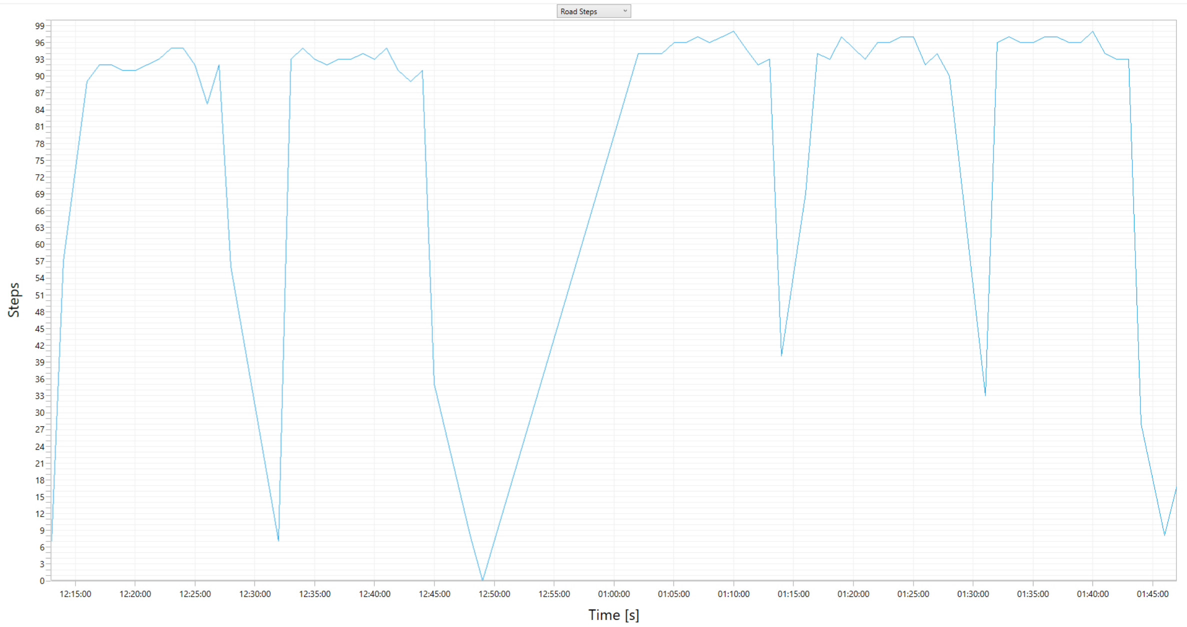


Figure 5.1: Chart showing step amount during each interval

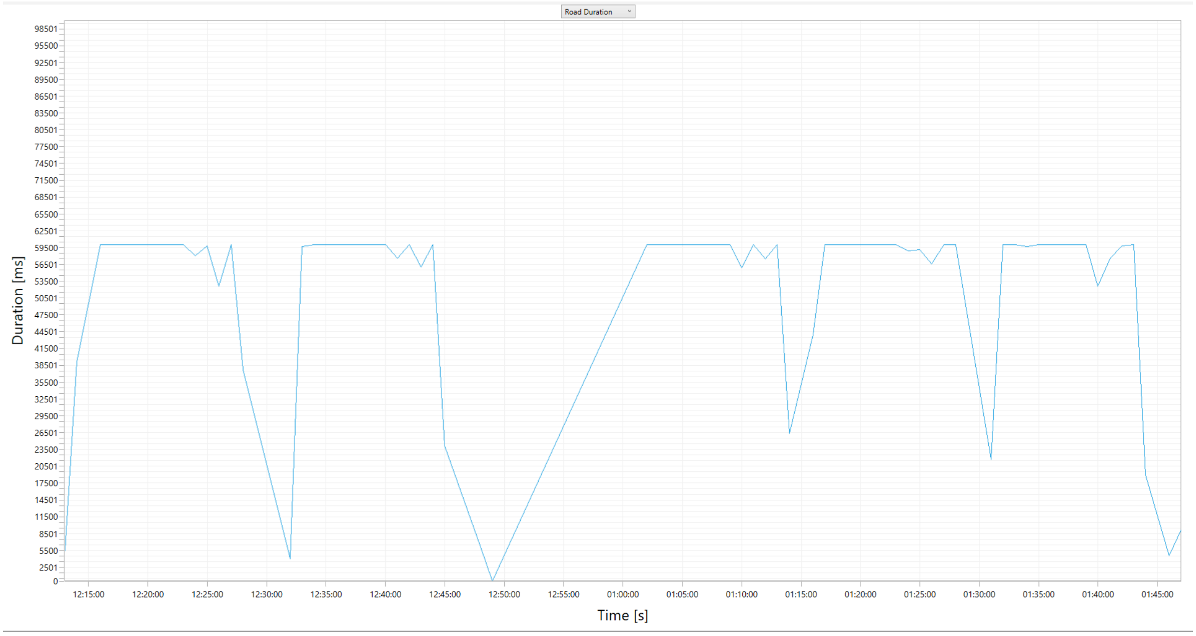


Figure 5.2: Chart showing the duration of each interval, which in general was a minute.

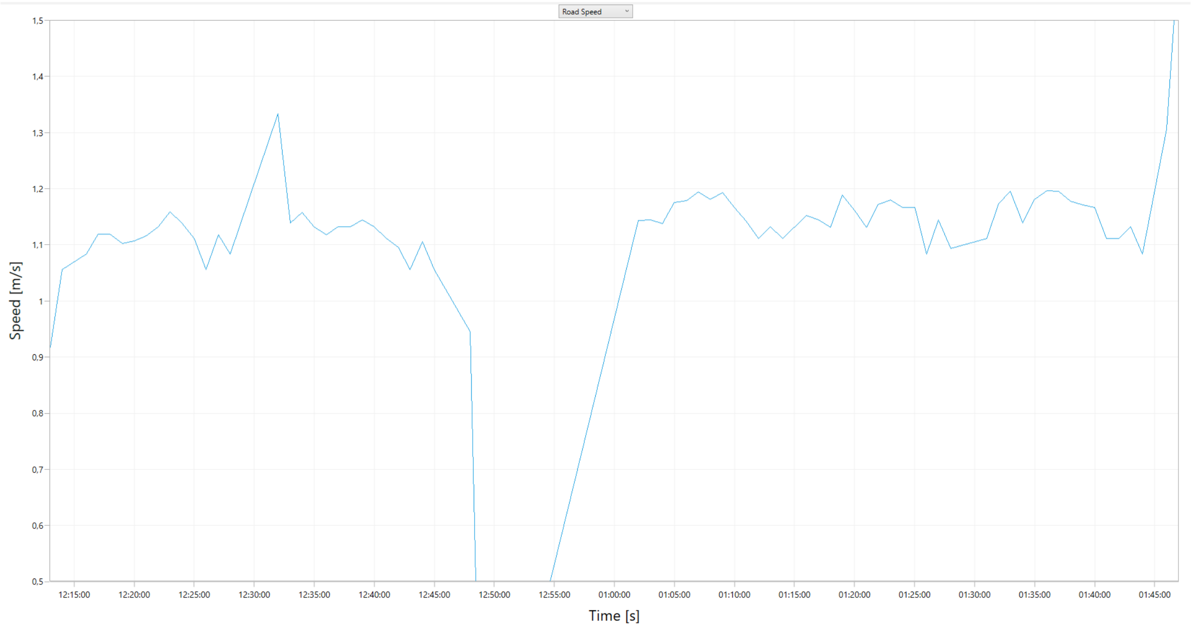


Figure 5.3: Chart showing walking speed in meters per second

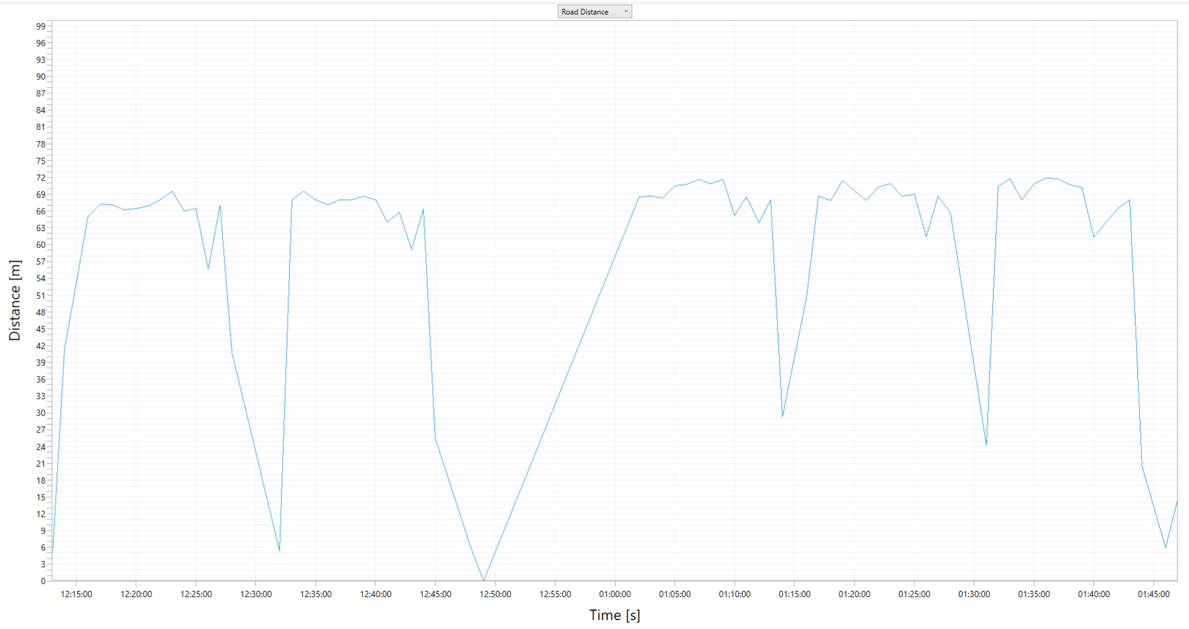


Figure 5.4: Chart showing distance traveled per interval

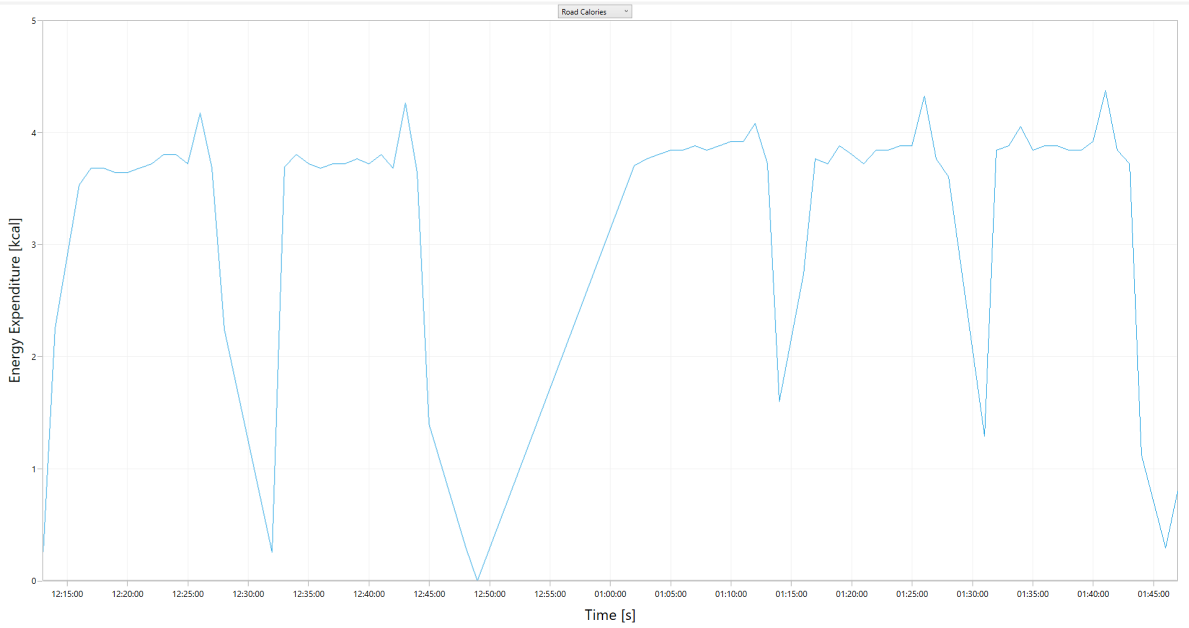


Figure 5.5: Chart showing amount of calories burned per interval

5.3 Answers from the interview

- Is the step chart intuitive and useful? Are there some improvements that could be made?
  - James suggested that the prototype could include an explanation for the sharp drops in the graphs. It was necessary to explain that it was due to deliberately pausing the measuring of footstep data in Samsung Health.
- Is the duration chart intuitive and useful? Are there some improvements that could be made?
  - James asked for an explanation for smaller drops in the chart that preceded the larger drops, theories were posited but ultimately it is unclear why they exist. He also seemed to be confused about what the duration meant and it had to be explained to him that it was the interval of time between each recording of data, which in general was every minute.
- Is the speed chart intuitive and useful? Are there some improvements that could be made?
  - James thought the graph made sense and couldn't think of further improvements that could be made. The graph had a single sharp drop which was explained by having to stop to adjust equipment and turning off Samsung Health measurement during that time.
- Is the distance chart intuitive and useful? Are there some improvements that could be made?
  - James thought it looked fine and matched previous graphs.
- Is the calorie chart intuitive and useful? Are there some improvements that could be made?
  - James thought this graph looked fine too.
- How can the overall presentation of the information be improved?
  - James had a suggestion that geo-coordinate data could be overlayed on the graphs to compare to. He also had a suggestion to add the metric system to the values (for example, the speed is

only showcased in meters per seconds in the speed graph). Lastly, James thought it would be interesting if the graphs also showed heart rate.

- How can footstep data be used as evidence in court?
  - Traditionally, according to James, footprints present on a crime scene would be studied to aid in the forensic investigation. The footprints could tell the forensic examiner whether the person was running or walking, or how the person walked. The footprints could also be measured to aid in identifying the perpetrator. However, footprints are unreliable as forensic evidence since they may be altered by for example weather. With the aid of digital tools, the forensic investigator is not limited to footprints as the only footstep data available to them.
- Can MSABs current technologies benefit from the addition of footstep metrics?
  - James saw that MSAB could benefit from footstep metrics. Not only could footstep metrics establish convicting evidence, but it could also establish evidence that someone is innocent.
- How can the prototype, after necessary improvements, be applied in law enforcement?
  - There are two main applications according James. The first one being to aid in initial investigatory measures. The second one being to present data in court. Being able to present complex data visually makes the data more intuitive and easy to understand, which is of great value.
- Are there any other comments?
  - James did not have any other comments.

Chapter 6

Discussion

The purpose of this chapter is to discuss and evaluate the results acquired and work done in the thesis. The chapter discusses and evaluates the research question, the method and future application.

6.1 Answering the research question

The research question defined in the introductory chapter is as follows: *How can footstep data be extracted from Android smartphones and visualized in a manner suitable for forensic analysis?* To answer the research question, this section will be divided into three parts. The first part will address whether the data can be used in court. The second part will address how this data can be extracted in a manner suitable for forensic analysis. The third part will address how this data can be visualized in a manner suitable for forensic analysis.

6.1.1 Footstep data as court evidence

The results from the comparison of accuracy between the pedometer and Samsung Health suggests that Samsung Health is more accurate than the pedometer. This also seems to be the case for longer distances, since footstep data recorded by Samsung Health in the larger experiments is more consistent with academic sources (such as [19] & [20]) than the pedometer. This suggests that the footstep data extracted from Samsung Health is accurate enough to be used as forensic evidence in court, since Samsung Health is more accurate than a commercial device created to measure step count.

To further support this claim, there are documented instances of data extracted from pedometers being used in court as evidence ([35], [36], [37], [38], [39], [40]). James Eichbaum also testified to this when speaking about how footstep data has traditionally been used as forensic evidence.

6.1.2 How to extract footstep data

As said in chapter 2, for artifacts to be considered viable as evidence, they have to satisfy some requirements [5]. The artifacts have to be:

- *Admissible*: Evidence must be gathered and preserved in such a way that it can be presented as evidence in court.
- *Authentic*: The evidence must be relevant to the case and have a clear and relevant origin.
- *Complete*: Evidence presented must be clear and complete in a way so that it shows the whole story.
- *Reliable*: Techniques used and evidence collected must not infringe on the authenticity of the evidence.
- *Believable*: The evidence presented must be clear, easy to understand and believable.

To achieve these requirements, the footstep data was extracted using two professional tools, XRY and XAMN. This allows the data to be extracted without human errors, such as overlooking data or accidentally changing artifacts, which could negatively impact the above mentioned requirements.

XRY and XAMN are also created to be used mainly by law enforcement agencies and military, which both have high standards on the digital evidence produced. To aid in evaluating the forensic tools, the evaluation reports written by the National Institute of Standards and Technology (NIST) was used [10].

6.1.3 How to visualize footstep data

When interviewing James Eichbaum, the authors asked how he could see the prototype being applied in law enforcement. In regards to data presentation, James said that being able to present complex information intuitively and in such a way that it becomes easy to understand is key. This is due to the fact that this information might need to be presented to non-technical people, such as judges or lawyers.

The authors chose to create a prototype containing graphs for each interesting set of data. Eichbaum approved this way of presenting data and said that it satisfied the above mentioned quality. Eichbaum did however suggest some improvements to the data presentation, such as better explanation of anomalies in the graphs, the inclusion of the metric system for the values, and lastly that more interesting data may be added, such as heartrate and geo-coordinates.

Because each of the problems addressed in the above sections have been solved, the hypothesis that the problem raised by the research question is solvable is confirmed to be true.

6.2 Evaluating method

This section evaluates the methods used to answer the research question. Specifically it will begin by evaluating the initial literature study. It will then discuss the generation and visualization of data using experiments and prototyping. Finally, answers from the interview with the expert are evaluated.

6.2.1 Literature study

The literature study resulted in a lot of context and knowledge being acquired, especially from the NIST testing report [10] and from the paper by Jan Peter van Zandwijk and Abdul Boztas [15]. The NIST study led to conversations with supervisors at MSAB about which bugs that NIST brought up had been solved. The paper introduced a background to how to investigate footstep data, albeit on a different platform.

However, there were issues due to mobile forensics being a niche field, resulting in the amount of sources that could be found to be sparse and some of those that were found could be a decade old.

The studies were in general fairly obscure and not often cited, which resulted in a somewhat hap-hazard method of searching for them, usually using Google services. In general there was also a lack of book sources.

6.2.2 Generating data

The method to conduct the experiments was changed several times. Initially workouts and public transport was to be included, but it was decided it wouldn't be relevant to the research question and was therefore done away with. The experiment method that was chosen served the purpose of generating data well and led to appropriate visualization.

The extraction went fairly smoothly and even the failed attempts at acquiring data were insightful into the structure of Samsung Health. The data found would be more useful for XRY LOGICAL than XRY PHYSICAL, since the method of extraction was through interacting with the app directly.

Converting the csv files that were acquired into a SQLite database was also fairly simple and the data was presented by DB Browser for SQLite in a fairly clear way. The method of extraction described should be able to serve as an instruction towards how to replicate it for readers even if they don't have in-depth forensic knowledge.

6.2.3 Visualizing data

Both the MSAB supervisor and the expert being interviewed were positive about the prototype and believed it presented information concisely and was optimized well. The creation of the prototype went without major issues, once necessary knowledge about how to use WPF, WPFChart and NuGet were acquired through online tutorials. However, learning did take some time and the authors created a test application before the actual prototype to get experience with the systems since previously they only had experience using Visual Forms, which is similar to WPF but slightly older.

Due to time constraints, the prototype lacks documentation, testing and bonus features, such as help pages and charts showcasing for example distance per step. These features were however discussed with supervisors previously and were seen as not being essential, especially since the point of a prototype is to create something in a timely manner that can be discarded in the future.

6.2.4 Interview

The interview was well prepared, with the questions being supervised before being sent to the expert before the interview. The interview itself went fairly smoothly, with the demo and questions being understandable to James. The conversation between the authors and James contained information about typical police procedures, which was very useful for determining the answers to the research question.

James also referred the authors to other police officers in case they had additional questions. But he was not taken up on the offer, because the interview seemed to have had sufficient information to answer the research question and preparing another interview would take too much time away from writing the report.

6.2.5 Validity and reliability of the method

The use of methodologies and data is for the purpose of answering the research question in a structured and coherent way. Therefore it is important for the methodologies and data to be valid, reliable and consistent. Without meeting these criteria the methodology and data would be irrelevant towards answering the research question, if not actively obscuring possible answers.

The validity of the literature study can be proven since sources were only evaluated if they met the 5 important criteria mentioned in chapter 2 [18]:

- *Purpose*: What is the author trying to accomplish?
- *Scope*: How much of the topic is covered and in what depth?
- *Authority*: What is the authority of the source – are the author and his or her credentials given?
- *Audience*: Who is the publication intended for?
- *Format*: How is the information presented?

Validity can be proven by the researcher demonstrating a clear cause and effect between elements. This is referred to as internal validity. Since the experiment consisted of manual data gathering and visualization, it was validated by the prototype creating visualizations that accurately portrayed the experiment [41].

Validity is also a measure of the accuracy of a proposition, usually supported by evidence. Answering the research question cannot be done with direct measurement, since the usefulness of our research to law enforcement agencies and mobile forensics developers cannot be reduced to numbers. The authors instead analyzed indicators to prove the validity of the results. This is known as construct validity and allows the use of expert opinions as an indicator that a specific interpretation of a measurement is valid [42].

The reliability of method is the consistency the method gives in its results [43]. For example, the same method under the same conditions should yield the same results, no matter who performs it. Reliability can be divided up into four types [44]:

- **Test-retest**: The level of consistency of a measure across multiple tests and time.
- **Interrater reliability**: The level of consistency of a measure when performed by multiple experimenters.
- **Parallel forms reliability**: The level of consistency of a measure for similar tests.
- **Internal consistency**: The level of consistency of each part of a measure in correlation to each other.

Out of these, test-retest has been the main way of proving the reliability in this thesis. It has been used mainly in the experiments by repeating the same test five times. Other parts of the reliability are also satisfied in the experiments. Parallel forms reliability can be seen when comparing the tests between the road and forest experiments. Both of them give approximately the same results. Internal consistency can be seen when looking at the average steps per minute, which are consistent. Another example of where one of the reliability types are satisfied is in the literature study where keywords were used. Similar keywords yielded similar results, which shows Parallel forms reliability.

6.3 Future work

This report showcases the ability for forensic tools to visualize footstep data retrieved from Android devices. Such research already existed for iPhone devices, and there are forensic tools that can extract data from Samsung Health on Android, but without visualizing the data.

As this research opportunity was advertised by MSAB, it shows that there is a demand by the forensic industry for these solutions. While the prototype is not in a state to be used commercially, both the supervisors and expert at MSAB agree that it serves as a proof-of-concept that it could be implemented in their proprietary technologies. It also details steps towards extracting footstep data and the structure of that data which is very interesting to them.

As Samsung Health is a very popular application that is already pre-installed on Samsung devices, and only requires a user to open the application once to start tracking your footsteps, it is very relevant for criminal prosecution due to how common the data will be among the general criminal population and its ability to debunk alibis.

The authors believe that footstep data and personal health data in general will become increasingly more important for the digital forensic, law enforcement and judicial industries. It will also be very important to present this data in a digestible form to a layman with no experience in forensics, such as someone serving as a jury.

This is why simple chart visualization for this data is important, especially since Android is a very popular mobile platform. Usability is a priority for MSAB, which is why it has invested in applications such as XRY EXPRESS which offers a simplistic user interface consisting of big, readable buttons.

The form of extraction that was used was not based on extracting data directly from the hardware. If laws concerning personal health data [45] change, then Samsung Health may remove the option to retrieve personal health data. This would require public research into solutions such as hooking [46] or ram dumps [47], such as shown by: [48][49]. Hijacking cloud tokens is also a possibility, such as shown by: [50]. Finally, you can retrieve much more information, such as coordinates and heart rate by extracting health data from Samsung Health sensors as shown in: [51]. The authors encourage researchers who are interested in building upon this subject matter to investigate these avenues.

References

[1] A. Mylonas, V. Meletiadis, B. Tsoumas, L. Mitrou, and D. Gritzalis, “Smartphone forensics: A proactive investigation scheme for evidence acquisition,” in *Information Security and Privacy Research: 27th IFIP TC 11 Information Security and Privacy Conference, SEC 2012, Heraklion, Crete, Greece, June 4-6, 2012. Proceedings 27*. Springer, 2012, pp. 249–260. [Pages 1 and 6.]

[2] “When police are hackers: Hundreds charged as encrypted network is broken - the new york times,” https://www.nytimes.com/2020/07/02/world/europe/encrypted-network-arrests-europe.html, (Accessed on 02/11/2023). [Page 1.]

[3] “About us - msab,” https://www.msab.com/about-us/, (Accessed on 02/13/2023). [Page 2.]

[4] J. Kävrestad, *”Fundamentals of Digital Forensics”*. Springer Nature, 2018. [Page 5.]

[5] R. Tamma, O. Skulkin, H. Mahalik, and S. Bommisetty, *Practical mobile forensics: Forensically investigate and analyze iOS, Android, and Windows 10 devices*. Packt Publishing Ltd, 2020. [Pages 6 and 40.]

[6] “privacy — dictionary.cambridge.org,” https://dictionary.cambridge.org/dictionary/english/privacy, [Accessed 21-Feb-2023]. [Page 6.]

[7] M. Losavio, P. Pastukov, and S. Polyakova, “Cyber black box/event data recorder: legal and ethical perspectives and challenges with digital forensics,” *Journal of Digital Forensics, Security and Law*, vol. 10, no. 4, p. 4, 2015. [Page 6.]

[8] “Xry_product_family_en.pdf,” https://www.msab.com/wp-content/uploads/2023/01/XRY_Product_Family_EN.pdf, (Accessed on 02/19/2023). [Pages 7 and 8.]

[9] “Xamn_family_en.pdf,” https://www.msab.com/wp-content/uploads/2023/01/XAMN_Family_EN.pdf, (Accessed on 02/18/2023). [Pages 8 and 9.]

[10] “test results for mobile device acquisition tool: xry office v10.0 build 10.000.268 – xamn v7.0.0 build 507_2022,” Apr 2022. [Online]. Available: https://www.dhs.gov/sites/default/files/2023-01/23_0127_st_test_results_for_mobile_device_final_mobile_forensics_v2021.09.28.pdf [Pages 9, 40, and 41.]

[11] “About nist,” Jan 2022. [Online]. Available: https://www.nist.gov/about-nist [Page 9.]

[12] R. F. A. P. A. H. O. N. Ademola O. Adesina, Kehinde K. Agbele, “Ensuring the security and privacy of information in mobile health-care communication systems,” http://www.scielo.org.za/pdf/sajs/v107n9-10/v107n9-10a12.pdf, (Accessed on 03/06/2023). [Page 10.]

[13] I. B. X. Z. Courtney Hassenfeldt, Shabana Baig, “”map my murder: A digital forensic study of mobile health and fitness”,” https://digitalcommons.newhaven.edu/electricalcomputerengineering-facpubs/94/, (Accessed on 03/06/2023). [Page 10.]

[14] A. Azfar, K.-K. R. Choo, and L. Liu, “Forensic taxonomy of popular android mhealth apps,” *arXiv preprint arXiv:1505.02905*, 2015. [Pages 11 and 12.]

[15] J. P. van Zandwijk and A. Boztas, “The iphone health app from a forensic perspective: can steps and distances registered during walking and running be used as digital evidence?” *Digital investigation*, vol. 28, pp. S126–S133, 2019. [Pages 11, 18, and 41.]

[16] S. Keele *et al.*, “Guidelines for performing systematic literature reviews in software engineering,” 2007. [Page 17.]

[17] S. Jalali and C. Wohlin, “Systematic literature studies: database searches vs. backward snowballing,” in *Proceedings of the ACM-IEEE international symposium on Empirical software engineering and measurement*, 2012, pp. 29–38. [Page 17.]

[18] R. B. Brown, *Doing Your Dissertation in Business and Management*. London: SAGE, 2006. [Pages 17 and 43.]

[19] D. R. Bassett Jr, H. R. Wyatt, H. Thompson, J. C. Peters, and J. O. Hill, "Pedometer-measured physical activity and health behaviors in united states adults," *Medicine and science in sports and exercise*, vol. 42, no. 10, p. 1819, 2010. [Pages 18 and 39.]

[20] J. Berko, R. Z. Goetzel, E. C. Roemer, K. Kent, and J. Marchibroda, "Results from the bipartisan policy center's ceo council physical activity challenge to american business," *Journal of occupational and environmental medicine*, vol. 58, no. 12, p. 1239, 2016. [Pages 18 and 39.]

[21] "Samsung health - apps on google play," accessed 25-Apr-2023. [Online]. Available: https://play.google.com/store/apps/details?id=com.sec.android.app.shealth [Page 18.]

[22] "What is samsung health? | samsung health tips | samsung uk," accessed 25-Apr-2023. [Online]. Available: https://www.samsung.com/uk/mobile-phone-buying-guide/samsung-health-tips/ [Page 18.]

[23] "What is test-retest reliability and why is it important? | cambridge cognition," https://www.cambridgecognition.com/blog/entry/what-is-test-retest-reliability-and-why-is-it-important, (Accessed on 03/17/2023). [Page 19.]

[24] "(9) the benefits of prototyping and validation | linkedin," https://www.linkedin.com/pulse/benefits-prototyping-validation-joeri-van-cauteren/, (Accessed on 03/17/2023). [Page 19.]

[25] "What is windows presentation foundation - wpf .net | microsoft learn," https://learn.microsoft.com/en-us/dotnet/desktop/wpf/overview/?view=netdesktop-7.0, (Accessed on 03/25/2023). [Page 19.]

[26] "Nuget gallery | home," https://www.nuget.org/, (Accessed on 04/25/2023). [Page 19.]

[27] "About wpf charts control | syncfusion," https://help.syncfusion.com/wpf/charts/overview, (Accessed on 04/25/2023). [Page 19.]

[28] "Patterns - wpf apps with the model-view-viewmodel design pattern | microsoft learn," https://learn.microsoft.com/en-us/archive/msdn-magazine/2009/february/patterns-wpf-apps-with-the-model-view-viewmodel-design-pattern, (Accessed on 04/25/2023). [Page 19.]

[29] "(pdf) assessing the reliability and validity of expert interviews," https://www.researchgate.net/publication/228664379_Assessing_the_Reliability_and_Validity_of_Expert_Interviews, (Accessed on 03/17/2023). [Page 20.]

[30] "Interviews - qualitative study design - libguides at deakin university," https://deakin.libguides.com/qualitative-study-designs/interviews, (Accessed on 04/14/2023). [Page 20.]

[31] "What is zoom? the popular video-chatting app explained," https://www.businessinsider.com/guides/tech/what-is-zoom-guide?r=US&IR=T, (Accessed on 03/25/2023). [Page 20.]

[32] "Top 4 data analysis techniques | maryville online," https://online.maryville.edu/blog/data-analysis-techniques/, (Accessed on 03/20/2023). [Page 21.]

[33] "Amidug/wpfalibaba," https://github.com/AmiDug/WPFalibaba, (Accessed on 04/25/2023). [Page 26.]

[34] "Interview with james eichbaum," https://drive.google.com/file/d/1xoywVNEXKxfIK6Ue1xp2s_bwlezF-v5y/view, (Accessed on 05/01/2023). [Page 29.]

[35] "State v. powasnik," p. 146, 1996. [Page 40.]

[36] "State v. jones," p. 782, 2007. [Page 40.]

[37] "State v. bedford," p. 758, 2003. [Page 40.]

[38] "In re nat. airlines, inc." p. 249, 1977. [Page 40.]

[39] "State v. acosta-diaz," 2004. [Page 40.]

[40] "Could your fitbit data be used in court?" [Accessed 08-May-2023]. [Online]. Available: https://www.pajcic.com/fitbit-data-used-court/ [Page 40.]

[41] "6. quintao et al," https://files.eric.ed.gov/fulltext/EJ1294617.pdf, (Accessed on 03/17/2023). [Page 43.]

[42] "Construct validity in software engineering | ieee journals & magazine | ieee xplore," https://ieeexplore.ieee.org/document/9780058, (Accessed on 03/17/2023). [Page 44.]

[43] F. Middleton, "Reliability vs. validity in research | difference, types and examples," 2023, [Accessed 15-May-2023]. [Online]. Available: https://www.scribbr.com/methodology/reliability-vs-validity/ [Page 44.]

[44] ——, "The 4 types of reliability | definitions, examples, methods," 2023, [Accessed 15-May-2023]. [Online]. Available: https://www.scribbr.com/methodology/types-of-reliability/ [Page 44.]

[45] "Privacy code of conduct on mobile health apps | shaping europe's digital future," https://digital-strategy.ec.europa.eu/en/policies/privacy-mobile-health-apps, (Accessed on 05/15/2023). [Page 45.]

[46] "What is hooking - definition of hooking | vmray," https://www.vmray.com/glossary/hooking/, (Accessed on 05/15/2023). [Page 45.]

[47] "Why ram dumping is so important and what tool to use?" https://belkasoft.com/ram-dumping-tool-selection, (Accessed on 05/15/2023). [Page 45.]

[48] "Decrypting databases using ram dump - health data - cellebrite," https://cellebrite.com/en/decrypting-databases-using-ram-dump-health-data/, (Accessed on 05/15/2023). [Page 45.]

[49] "Android forensics, smart flow, selective file system extraction – part 2 of cellebrite solutions 2022 update summary - cellebrite," https://cellebrite.com/en/android-forensics-smart-flow-selective-file-system-extraction-part-2-of-cellebrite-solutions-2022-update-summary/, (Accessed on 05/15/2023). [Page 45.]

[50] "Samsung device data extraction in oxygen forensic® detective - forensic focus," https://www.forensicfocus.com/news/samsung-device-data-extraction-in-oxygen-forensic-detective/, (Accessed on 05/15/2023). [Page 45.]

[51] "Extract health data from your samsung device | by anuradha wickramarachchi | towards data science," https://towardsdatascience.com/extract-health-data-from-your-samsung-96b8a2e31978, (Accessed on 05/15/2023). [Page 45.]

€€€€ For DIVA €€€€

{
”Author1”: { ”Last name”: ”Dugiev”,
”First name”: ”Amiran”,
”Local User Id”: ”u1sqvc8q”,
”E-mail”: ”amiran@kth.se”,
”organisation”: {”L1”: ”School of Electrical Engineering and Computer Science”,
}
},
”Author2”: { ”Last name”: ”Cassé”,
”First name”: ”Henrik”,
”Local User Id”: ”u1pgcu0j”,
”E-mail”: ”casse@kth.se”,
”organisation”: {”L1”: ”School of Electrical Engineering and Computer Science”,
}
},
”Cycle”: ”1”,
”Course code”: ”II142X”,
”Credits”: ”15.0”,
”Degree1”: {”Educational program”: ”Degree Programme in Computer Engineering”
,”programcode”: ”TIDAB”
,”Degree”: ”Bachelors degree”
,”subjectArea”: ”Technology”
},
”Title”: {
”Main title”: ”Forensic Analysis of Footstep Data”,
”Language”: ”eng” },
”Alternative title”: {
”Main title”: ”Forensisk analys av fotsteg data”,
”Language”: ”swe”
},
”Supervisor1”: { ”Last name”: ”Lindbäck”,
”First name”: ”Leif”,
”Local User Id”: ”u1w0ljrk”,
”E-mail”: ”leifl@kth.se”,
”organisation”: {”L1”: ””,
”L2”: ”Computer Science” }
},
”Supervisor2”: { ”Last name”: ”Svedman”,
”First name”: ”Kjell”,
”E-mail”: ”Kjell.Svedman@msab.com”,
”Other organisation”: ”MSAB”
},
”Supervisor3”: { ”Last name”: ”Warnicke”,
”First name”: ”Emil”,
”E-mail”: ”Emil.Warnicke@msab.com”,
},
”Examiner1”: { ”Last name”: ”Galjic”,
”First name”: ”Fadil”,
”Local User Id”: ”u1hkgpgs”,
”E-mail”: ”fadil@kth.se”,
”organisation”: {”L1”: ””,
”L2”: ”Computer Science” }
},
”Cooperation”: { ”Partner_name”: ”MSAB”},
”National Subject Categories”: ”10201, 10205, 10206”,
”Other information”: {”Year”: ”2023”, ”Number of pages”: ”1,51”},
”Copyrightleft”: ”None”,
”Series”: { ”Title of series”: ”TRITA-EECS-EX” , ”No. in series”: ”2022:00” },
”Opponents”: { ”Name”: ”A. B. Normal & A. X. E. Normalè”},
”Presentation”: { ”Date”: ”2022-03-15 13:00”
,”Language”:”eng”
,”Room”: ”via Zoom https://kth-se.zoom.us/j/ddddddddddd”
,”Address”: ”Isafjordsgatan 22 (Kistagången 16)”
,”City”: ”Stockholm” },
”Number of lang instances”: ”2”,
”Abstract[eng]”: €€€€

```
\noindent
Digital forensics is a niche field of study that encompasses such things as extraction, analysis and
presentation of digital information, that could be used to produce forensic evidence. There are
several companies whose sole specialization is providing technical software solutions to detectives
to help them quickly analyze retrieved devices that might contain evidence, instead of having to send
the devices to forensic labs. \\ \\
However, there are still many areas that aren't fully explored within the digital forensics industry,
such as using personal health data. A factor that confounds this problem is that there are many
different mobile devices running different operating systems and different versions of different
```

applications. This report examines footstep data extraction and visualization on Android. Whether this data could be used as evidence according to law enforcement agencies is also investigated. \\ \\
Following a literature study to gain knowledge of the field of digital forensics, an experiment was conducted to gather data on a device through the Samsung Health application. This data was extracted and converted into a database, which was visualized using a prototype in the form of charts. Finally a technical trainer and former police officer was interviewed regarding whether the prototype could be seen as a proof-of-concept for future implementation among digital forensics solution providers. It was concluded that step data visualized in the form of graphs would be useful as forensic evidence for law enforcement detectives and juries.

€€€€,
”Keywords[eng]”: €€€€
Digital forensics, Samsung Health, Footstep data, Data visualization, Android €€€€,
”Abstract[swe]”: €€€€

\noindent
Digital forensik är ett studieområde som omfattar extraktion, analys och presentation av digital information, som kan användas för att producera forensiskt bevis. Det finns flera firmor som specialiserar i att utrusta detektiver med mjukvarulösningar som kan analysera enheter som kan innehålla bevis, istället för att skicka enheten till ett forensiskt lab. \\ \\
Det finns dock många områden som inte är helt utforskade inom digital forensik, som
användning av personlig hälsodata. En faktor som utökar detta problem är att det finns många olika mobila enheter som kör olika operativsystem med olika versioner av olika applikationer. Denna rapport undersöker fotstegsdata extraktion och visualisering på Android. Om denna data kan användas som bevis av rättsväsende blir också utforskat. \\ \\
Efter en litteraturstudie för att få mer kunskap inom digital forensik, utfördes ett experiment för att samla data på en enhet genom Samsung Health-applikationen. Detta data extraherades och konverterades till en databas, som visualiserades genom en prototyp i form av grafer. Slutligen intervjuades en teknisk tränare och ex-polis för att se om prototypen skulle kunna ses som ett bevis på att digital forensik skulle kunna implementera stöd för fotstegsdata i framtiden. Slutsatsen drogs att stegsdata visualiserat i form av grafer skulle vara användbara som forensiskt bevis för detektiver och juryer.

€€€€,
”Keywords[swe]”: €€€€
Digital forensik, Samsung Health, Fotstegsdata, Data visualisering, Android €€€€,
}