School of Computing and Communications
Lancaster University

A Framework for Privacy Aware Design in Future Mobile Applications

Sarah Mukisa
B.Sc, M.Sc.

Submitted for the degree of
Doctor of Philosophy
July 2019

Abstract

Mobile communications and applications play an important role in connecting people ubiquitously across different domain spaces due to their portable nature and easy accessibility. Mobile applications have drastically changed the way businesses are run by bringing them closer to their customers. Businesses today are connected to cloud based-tools, which makes it easier to start and run a business. Furthermore, mobile applications have changed the way we communicate with each other in our daily lives. They have increasingly been deployed by companies to help with, among other things, the management of business efficiency, ease in accessing information, simplifying communication and the provision of user-friendly applications. The number of mobile devices is increasing exponentially, it is estimated that 1.5 billion devices are available to the public worldwide. In addition, there is a multitude of operating systems running on these devices, all running on different architectures and configurations. The diversity of the different versions of applications that need to be constantly updated as they become outdated makes mobile applications highly susceptible to security and privacy flaws. Until recently, privacy has not been the main centre of interest within the design of mobile applications. Although, a number of privacy preserving solutions have been developed to improve privacy, existing research solutions adopt static design models which are not suitable for mobile applications. There is a significant gap between having common practices for designing and implementing privacy-preserving methods due to the cross-disciplinary nature of mobile applications. Most importantly, personal data are constantly collected and shared with unknown recipients. This is a challenging problem as users are not aware of how their data is used and shared without their consent. Furthermore, existing privacy policies are not stringently

implemented during application development. Application designers do not comply with regulations envisaged by data protection regulation bodies. To investigate the problem domain, this thesis takes a bottom-up approach and contributes by analyzing current mobile applications to determine the integration of privacy mechanisms and privacy policies at the application level. We should however note that, the focus of this work contributes to the knowledge related to designing of holistic privacy preserving mobile applications and not the implementation aspect. Furthermore, this thesis introduces a novel privacy trade-off analysis framework that enables the design of privacy-aware applications. A privacy trade-off analysis generates a design solution that best suits an application's privacy goals and requirements. To demonstrate the privacy-aware framework, TRANK, two prototypes in the eHealth domain and the V2X Telematics domain, that integrate privacy-preserving technologies in modern mobile applications have been implemented and tested. Our implementation takes into consideration the trade-off between privacy, functionality and performance to provide a better privacy-aware application. The resulting system enables users to choose which data are to be collected about them. In this way, users can easily opt in and out of the application without having to give up all their personally identifiable information whenever they choose to, thus, enhance their overall privacy preservation. To the best of our knowledge our framework and the results in this thesis out perform the existing state of-the-art privacy preserving solutions. The privacy-enhancing technologies employed and the privacy-by-design mechanisms introduced at the initial stages of development thus, aid the improvement of privacy in mobile applications.

Dedication

Success is getting what you want. Happiness is wanting what you get.

Dale Carnegie.

To my grandparents, who sowed the seed of education in us.

To my family. Thank you very much for the patience.

Acknowledgement

I would like to thank all those who have helped me to achieve my goal of completing a PhD. The process has been a long and tedious one.

First of all I would like to thank my supervisors; I would like to thank my supervisor Dr. Angelos Marnerides for his continuous support and constructive guidance. You provided me with the tools that I needed to choose the right direction and successfully complete my dissertation. Thank you for believing in me and enduring my endless emails and conversations regarding this work.

I would like to thank Professor Awais Rashid, whose expertise was invaluable in the formulating of the research topic and methodology. Thank you for constantly assuring me that I would succeed.

I would also like to thank my lecturer Dr. Steffen Zschaler at Kings College London, who initiated the process of my PhD studies.

Thank you very much Hauke Sommerfeld for the support you gave during this thesis.

Finally I would like to thank my family for their continuous support.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Mobile applications are used for a range of online services across private and commercial domains. These domains need to be secured and resilient to face challenges of privacy leakages, privacy misconfigurations, cyber security attacks and system failures. Current mobile applications have, however, faced a lot of criticism about unauthorized and unintentional transfer of sensitive data due to misconfigured back ends, data sharing and transfer to third-party service providers. In particular, the range of beneficial opportunities and functional properties offered by mobile applications such as communication through OSNs, easy accessibility, better functionality, business enhancement, introduce a number of privacy and security vulnerabilities. An indirect drawback lies within mobile applications dependency on the Internet where privacy protection has been extensively studied but still faces setbacks.

Data privacy in mobile applications significantly depends on how applications are designed and implemented. Mobile applications generally operate using operating systems that are installed on handheld devices (e.g. smart-phones, tablets, wearables, smart-watches e.t.c) with a plethora of programs running on them. The handheld devices employ various sensors ranging from geo-location sensors (and other sensors e.g., cameras, microphones, accelerometers, motion sensors) to heartbeat sensors as depicted in eHealth

applications. These applications acquire a lot of personal and sensitive data from their users. This data is often transmitted to large service providers and third parties without users' knowledge leading to a variety of data privacy concerns. Such type of data can easily be intercepted if not securely transmitted and stored leading to massive data breaches. In addition, due to the different applications that run on these devices e.g. geo-location applications, users are easily tracked based on their daily movements. Furthermore, an immense amount of data is transmitted to online data collection companies e.g., OSN companies and users have been subjected to total surveillance [Bat19] Most of the applications are connected to the Internet, with a constant connection to various applications like search engines, geolocation applications and Online Social Networks (OSN) companies like data brokers and third-party service providers use the collected data to provide targeted ads. Consequently, there has been a surge of privacy data breaches in recent years due to the huge volumes of data collected and the high numbers of users connected to such networks. This thesis aims at enhancing privacy preservation in current and future mobile applications by introducing a privacy design framework which employs privacy-by-design through the whole application development lifecycle.

We introduce a trade-off analysis framework that aids application designers in considering the tradeoff between privacy and functionality during system development in an attempt to deliver privacy aware applications that minimize data collection while maintaining application functionality. This way privacy challenges in current mobile applications can be solved to protect user privacy.

1.2 Aims

Future mobile applications must be designed in such a way that privacy enhancing mechanisms are integrated at the go and are pro-active and not only re-active to any data breaches that occur. Users should be aware of the data they provide to data processing companies and should be able to consent to which data they provide. The aim of this thesis therefore is to:

- Create user awareness about the data being collected in mobile applications.

- Create a framework that designs privacy-aware applications to reduce the amount of personal data collected by mobile applications.
- Aid application designers in decision making about privacy preservation in the early stages of system development.
- Aid developers improve the privacy-aware design of mobile applications.
- To expand and contribute towards transferable knowledge in the context of privacy and particularly for data privacy in mobile applications

For the above aims to be achieved, the main technical objective of this thesis is to present a framework that aids application developers in designing privacy-aware applications at the initial stages of system development. In the following, we elaborate on some of the challenges and privacy breaches that have manifested in modern mobile applications.

1.3 Motivation

Mobile communication systems are composed of voice and multimedia data transmitted over wireless communication technologies. In recent years, we have seen a surge of privacy and security attacks in mobile communications leading to both financial and personal information loss. Mobile apps are the main focus of this research because of their evolving nature. Due to their rapid development they require special attention because they run on multiple platforms that are often hard to operate on various devices. This pervasive property of mobile applications therefore, makes the development of privacy and security across the multi platforms very challenging.

Although organizations are constantly improving their systems in terms of privacy preservation, data breaches have continued to occur. This may occur when privacy attacks are caused by a third party or through breaches of privacy trust by data collectors. Privacy attacks caused by third parties are illegal whereas the breach of privacy trust by data collectors is permissible under the terms and conditions of the agreement with the

company. This is usually included in privacy policies and often read by users or generally accepted without knowing the consequences.

We consider both aspects of data breaches similarly as they all lead to damage and privacy abuse. The cost of damages caused increases continuously. It is estimated that the average cost of a data breach in 2017 was 3.5 million dollars. This was explicitly depicted by the drop of Yahoos market price by 350 million dollars when Verizon bought it in 2017 after it was reported that, 3bn records had been stolen by hackers [2].

The records included personally identifiable information such as names, telephone numbers, dates of birth, email addresses, hashed passwords and security questions and answers to the questions. The breach was done by using false cookies which gained access to users accounts without using a password. Hackers were able to impersonate themselves as the owners of the email accounts leading to massive data theft. To date, this is regarded as the worst data breach in history [3]. The past years have proved that online data we assume to be in safe hands is not as safe as we thought it to be. We have seen massive record theft from big companies ranging from Online Social Networks to business and medical bodies. In the following we elaborate on major data breaches that have occured in recent years.

Online Social Networks: Online Social Networks(OSN) have increased in the last decade e.g., Facebook, Myspace, Twitter, Instagram, Snapchat which have a large number of users, however, these have been targeted as they do collect a lot of personal data. The interconnection to friends profiles' allows hackers to extract large amounts of data from the network. One of the major OSN data breaches was hit by Facebook in 2018 when 50 million user accounts were harvested by Cambridge Analytica a data analytics firm and used to manipulate voters during the 2016 US election campaigns and the Brexit vote [4]. This caused a market share price drop of 3% on the stock market [5].

In another OSN incident, Myspace reported that approximately 427 million records were leaked accessing usernames, passwords, and email addresses in 2016. Although Myspace is considered as a dormant OSN, unfortunately, user data is never deleted and data breaches may occur without users actively using the OSN. Not only do data breaches

occur in large data housing companies, in addition to this some companies do not disclose the data breaches to their customers. Google+ exposed 497,000 data records in 2018, which consisted of full names, birth dates, profile photos, places lived, occupation and relationship status. The data breach enabled third-party applications to pull data from users and their friends. As a result, all user functionality was closed down. Nevertheless, Google+ chose to cover up the data breach to avoid any regulatory scrutiny.

Medical data breaches: In the medical sector, it is estimated that securing healthcare systems will gradually increase and exceed 65B [6] in an effort to avoid being attacked. Medicaid, one of the Centers for Medicare and Medic agencies in the USA has been a major target for attacks. In 2017, Molina a Medicaid insurer, shut down its patient portal after a major data leak exposed patients records. A total of 4.8 million records were compromised which included, diagnosis, medication, dates of birth, names, and addresses. This follows that, whoever, had access to these records is able to identify the patients' addresses and their diagnosis [7].

In a similar incident, 21st Century Oncology a cancer healthcare provider based in Florida was targeted by hackers who stole 2.2 million patient records. The records included names, physician names, treatment plans, social security numbers, and insurance data [8]. In May 2018, UnityPoint Health based in Iowa reported a data breach caused by an email phishing attack which affected 1.4 million records. This is so far the largest medical data breach in 2018. Emails accounts were compromised and were responded to by staff therein giving away a lot of sensitive data to the hackers. MyHeritage genealogy site was hacked in 2017 and 92 million user accounts which included email addresses and hashed passwords. The compromised data was found on a private server under a file name called "MyHeritage". Although MyHeritage acted fast in informing its users, the data set can still be used for illegal purposes. Not only private companies have been a focus of such data breaches, but government bodies have been targeted as was seen in Singapore. This attack led to 1.5 million patients being compromised including those of the prime minister from the Singapore government's health database [9].

Interestingly, some of these attacks are done by well-known hacker groups who are

hard to stop and prosecute. In the UK, a group called Dark Overload comprised celebrity and Royal family data records from a prominent surgery facility called the London Bridge Plastic Surgery Clinic. Intimate health records such as plastic surgery before and after images, genital modification images, breast enhancement information were sent to the media to prove that they had stolen the data [10].

What is worrying about the scale of health data theft is that patients suffer from data breach consequences and the physical pain of the illnesses they are diagnosed with, which are sometimes terminal. In addition, health care data is permanent and cannot be changed like in the case of passwords or credit card numbers. This explains why medical data is sold at a higher price on untraceable black market websites like the Darkweb.

Dating sites: Dating in the 21st century has changed in various ways but the most significant of all are the mushrooming mobile dating applications online e.g. Bumble, Happn, Hinge, Badoo, Tinder, Grindr, Match, eharmony to mention but a few. These dating sites collect a lot of intimate data used to match users with best-qualifying partners. Data collected includes geo-location data for matching users with partners in the neighborhood, pictures, interests, particulars of partners one is interested in, Facebook IDs, names, email addresses, facebook likes e.t.c. This type of personal data easily attracts hackers and if not properly secured can be leaked to the public as intruders are constantly searching for loopholes to compromise the networks.

In 2014, Tinder a major mobile dating app exposed the geo-locations, Facebook IDS, gender, birthdates, and username of its customers. Tinder connects to Facebook to mine data from users and their friends' profiles. This interconnection makes personal data susceptible to a lot of data breaches that may occur both in Facebook accounts as well as Tinder accounts.

In another incident, the gay dating app Grindr was reported to have exposed the HIV status and geolocation data of its customers to third party companies. Its surprising to know how much private information users are willing to submit on dating apps. For example, Tinder asks for the details of HIV prevention and exact HIV status which in turn is used to match with others. Details of the medication users are on, or if the viral

load is low and HIV can't be transmitted are also required. This means that whoever got access to this data could identify the exact location and HIV status and medication used by the individual [11].

Fitness application data breaches: The emergence of eHealth apps or mobile health apps has created new revenue generating opportunities for medical companies in fields like diabetic management by enabling consumers to monitor their health status continuously. These apps are a big relief to customers and offer very good services towards health care management. However, current medical apps have continuously collected personal and sensitive data without traditional scrutiny from medical bodies. The most used eHealth apps are fitness apps. Several data breaches have been reported in fitness apps. The Under Amour data breach from its fitness nutrition app MyFitnesspal exposed 150 million data records. The app however exposed partial information of the dataset comprising of usernames, passwords, and addresses. Other identifiable information like location information and credit card numbers was not leaked thus reducing the damage caused [12].

Another fitness app that has reported data breaches is Pumpup, which exposed 6 million records of user data like health data, emails, credit cards, location data, facebook accounts, private conversations. The data sets were stored on Amazon cloud infrastructure through an insecure server [13].

In November 2017, Strava a fitness tracking app released a heat map which shows the activity of its users online. This, however, raised a lot of privacy concerns. The app was able to expose the activity of military personnel in military bases of war tone areas like Afghanistan. Thus, endangering personnel in case of a counter-attack from terrorists [14].

Most of the fitness apps are connected to Online Social Networks (OSNs) which are used to register users. A good example of users' unawareness is that of a Reddit user. He asked for advice from Fitbit users about his wife's device which was faulty due to changes in the Fitbit tracking measurements. He was surprised to be informed that his wife might be pregnant which was revealed to be true [15]. This indicates that users are not well informed about the implications of the data they provide to app companies.

Privacy concerns also emerge from data collection done in weightloss apps which provide solutions to weight loss using genetic data. EnbodyDNA a product of the weightloss app Lose it!, promises weight loss by determining the gene responsible for losing weight. However, the field of personal genetics is unregulated and users do not understand who owns the DNA information. The gene information is sent by post without any information about who has access to it. It is reported that most genetic testing companies sell user data to third party service providers although users assume its protected [16].

Business domain data breaches: In the business entity, several companies e.g. E-commerce companies have been hacked into worldwide, for example, banks, online businesses, credit card companies, ticket vending companies, airline companies, telecommunication companies e.t.c. E-Business and E-commerce applications have exponentially increased in past years. We have seen a rise in Online shops of 80% for internet users in the UK alone with key major players like Amazon, eBay, Alibaba, Rakuten, Etsy, Craigslist, eBid. These applications require personal private data in order to do Business-to-Business (B2B) and Business-to-Customer (B2C) transactions.

Some of the data collected include credit card numbers, products bought, email addresses, delivery, and payment addresses e.t.c. Companies like Amazon have increased their advertising potential by using initiatives like amazon prime which also has led to the increase in customers and thus buying online. The gradual increase in the number of online customers and online shopping has led to many of the small retail stores on the high-street closing down.

However, apart from this concern, users are getting troubled about the data being collected while shopping online. The information collected is used to tailor users preferences according to details like credit card numbers, postcodes, products bought online e.t.c. The profiling is used to send tailored adverts to users' websites and mailboxes. This is mainly done by third parties. It is not yet clear how large data collection companies on the market deal with user data. These companies sell user data like browsing history, online shop email accounts, credit card history, marital status, postcodes to third-party providers and data brokers for the purpose of advertising. Concerns over the trading of

credit card details on the black market and on illegal sites like Dark Web for as cheap as £11 are increasing [17].

The credit monitoring company Equifax faced one of the major data breaches that exposed a lot of information. Hackers compromised 146.6 million records including social security numbers, credit card numbers, birthdates, drivers license numbers, passports. This is half of the US population. This leaked information may affect customers who need verification from financial creditors to acquire loans once the data gets manipulated [18]. Another major e-business domain worth mentioning is that of online banking. More banks are encouraging customers to sign onto online banking. Customers download online banking apps on their mobile phones regularly and banks encourage customers to sign onto online banking.

Although mobile banking apps are secured, fingerprinting methods can be used to track user behavior online using features like the operating systems, email addresses, gadget type, screen size e.t.c. Certain plugins are able to track user behavior between different browsers, for example, some banking institutions are known for using cookies that could potentially distinguish where a customer has a second bank account based on the browser information and activity.

Most users are not aware of this and are constantly monitored without their permission and knowledge. Therefore, users have to take caution in what type of information they provide in online banking and how they do online transactions keeping in mind the data privacy preferences.

The banking sector has been actively targeted, as was reported in May of 2018 by two large Canadian banks Simpli Financial and Bank of Montreal. The two banks were hacked and its estimated that a total of 90,000 clients lost their data which included personal and account information [19].

Airline companies have also been a target of online hackers. In the summer of 2018, British Airways was hit by a data breach which saw 380,000 transactions made through BA.com while making bookings and changes compromised. Holiday makers were disrupted by the data breach as they were advised by BA to contact their banks or credit card,

providers. Data stolen included email addresses, names, credit card numbers, CVV numbers found at the back of the credit cards and expiration dates. Customers were advised to either close their credit card accounts or monitor transactions on their accounts closely.

In general, there is a lack of privacy regulation and controls while designing mobile applications as we have seen in the above-mentioned data breaches. In for example, the eHealth domain, medical apps have been introduced to aid in patients analysis and disease management. The major concern though is that many of these apps do not go through stringent traditional quality controls and scrutiny from medical organizations, although they do collect a lot of private and personal sensitive data.

Similarly, modern apps e.g. vehicular applications are composed of inbuilt sensors which are capable of tracking location, social behavior and monitoring of day to day activities. The collected data is used to infer personal and social movements thereby generating a lot of privacy concerns. In addition to the sensor data, Vehicle-to-Everything (V2X) systems have introduced apps to offer services like telematics insurance to their customers. However, no limitations are put in using geolocation data.

Telematics insurance apps: Another emerging automotive technology is that of telematics insurance. Telematics insurance applications are on demand as they provide cheaper alternative insurance policies compared to the traditional car insurance policies. Privacy concerns have arisen as users are of the opinion that telematics insurance companies are constantly tracking and monitoring them. Additionally, drivers have concerns about their data being sold to third-party service providers without their consent. In this thesis, we focus on telematics mobile applications and investigate the privacy concerns related to using mobile applications.

In practice, apps have to follow stringent privacy requirements by providing privacy policies and open privacy practices. However, a major privacy concern that arises is whether current apps depict what is in the privacy policies they provide to their users? It is reported that privacy policies of tech giants like Yahoo, Facebook and Google are still not GDPR compliant much as the GDPR was enforced in May 2018.

Due to the data breaches named above, privacy awareness has become a major concern in recent years as companies continue to collect data without users consent. There is continuous monitoring of users through third-party companies which are outsourced to provide services to Online Social Networks (OSN).

User's location and demographic location have been used to profile users as has been seen by the Cambridge Analytica data scandal [20]. In this work, we investigate current privacy breaches in mobile apps and aim in finding solutions to existing privacy challenges faced in designing mobile applications.

1.4 Problem statement

Mobile applications and smart devices have become household items for the majority of the population. It is estimated that more than 60% of the world's population owns a mobile phone. Consequently, there has been a multitude of mobile applications developed and uploaded on to the Android market known as the Google Play store. However, these mobile applications come with privacy-related challenges. Android applications may reveal private data to both third-party applications or app developers.

Android apps use permissions to implement security and privacy. Much as the permissions are well structured in restricting the apps not to access user's data, some apps leak personal data to third-party service providers [21]. Furthermore, users are not aware of how this data is used by app designers and third party applications.

Users are advised to authorize third-party service providers before their personal data is shared as envisaged by the EU 2016/679 General Data Protection Regulation (GDPR). Such privacy preserving measures have been enacted by the EU legislative bodies. However, data brokers and third-party service providers continue to share and sell customers personal and private data [22]. Users are not explicitly told of the data and the implications of this data being collected.

The key problems of current mobile application development considered in the scope of this research therefore are:

- A lack of privacy awareness and user empowerment in mobile applications

- A lack of transparency from the mobile data companies that collect and store the data and subsequently share it with third parties without users consent.
- Lack of privacy policy implementation and control both at the application level and requirements and goals development level.
- Improper application design which does not integrate privacy enhancing technologies and privacy-by-design methodologies at the initial stages of system development

There is a significant gap in the way applications are designed and how users data is transmitted, stored and managed. The above incidences highlight the need for the design of less privacy-invasive applications. In this work, we address these challenges by investigating data collection, privacy policies, and privacy control measures to get an insight into the design and privacy implementations employed in current mobile applications. The main purpose of our study is to provide a framework named TRANK, for designing privacy-aware mobile applications. Through the TRANK framework we will expand and contribute towards practical knowledge on designing and implementing future mobile applications.

In detail, the framework;

- Defines privacy, functional and performance goals and designs privacy requirements using the privacy-by-design notion of the planned application.
- Reviews and redesigns conflicting goals and requirements to meet privacy preserving standards.
- Performs a functionality, privacy and performance trade-off analysis.

1.5 Thesis Overview

This thesis is structured as follows: In chapter 2 we present the background and related work of the research we have done in this thesis. Chapter 3 elaborates on the research methodology and provides the research methods used. Chapter 4 presents the contributions presented in this study which involve two major empirical studies based on two case

studies; the eHealth case study and a telematics case study used to investigate current mobile applications on the market. Chapter 5 presents a privacy aware trade-off analysis framework (TRANK) which is the core of this thesis. In chapter 6 we provide the technical implementation of TRANK in form of two mobile applications that act as a proof of concept for the framework and further presents an evaluation of the research done in this thesis considering the state-of-art of current mobile applications and the design of privacy aware applications. Finally chapter 7 gives conclusions to this thesis and future prospects that need to be considered while building privacy-by-design mobile applications.

1.6 Summary

This chapter provides an introduction to this thesis which attempts to address privacy and security challenges of current mobile This chapter provides an introduction to this thesis which attempts to address the privacy and security challenges of current mobile applications. The motivation section identifies our focus on the privacy-related data breaches that have occurred in mobile applications. The data breaches have been reported both in the media and by regulatory bodies. We elaborate that all technology-based domains are affected by privacy and security flaws that influence both user privacy and the way businesses are done. In particular, we have discussed the major breaches that have affected user privacy and security which are caused by systems not being properly designed and secured. This section presents an extensive discussion of the data breaches and emphasizes that measures have to be taken to address these issues in an attempt to reduce the data breaches. In the Problem Statement section, we identify the major problems faced through the use of mobile applications followed by the Aims and Objectives of this thesis. Next, we present a Thesis Structure of this work and finally we conclude with a brief summary of this chapter.

Chapter 2

Background and related work

This chapter presents background information on privacy-preserving mechanisms applied in general privacy research, which we think plays a great role in designing mobile applications.
Privacy research: One of the most influential definitions of privacy in previous research is that of legal scholar Alan Westin. He defines privacy as: "the claim of individuals, groups or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" [23]. Considering this definition, we need to understand how privacy is achieved. Privacy has been extensively discussed in previous research because individuals are becoming increasingly concerned about their personal data. Researchers have come up with various approaches to address this, such as the use of Public Key Infrastructures (PKI) used for encryption [24], firewalls [25], intrusion-detection systems [26], privacy-by-design [27](that is privacy-by-architecture and privacy-by-policy) approaches or privacy-enhancing technologies (PET) [28] to ensure privacy.

Although these approaches have been commonly adopted, experience has shown that IT infrastructure systems cannot be protected by such defences alone. We have seen significant numbers of privacy breach incidents in social media, credit card corporations etc. Social media companies are attacked repeatedly; for example, in 2013, nearly 2 million accounts' user names and passwords were reported stolen from social media giants Facebook, Google, Yahoo and LinkedIn, and other websites. Similarly, almost 40 million

credit cards' and debit cards' personal identification numbers (PINs) were stolen from the second-largest American retail corporation, Target [29]. Such privacy breaches have led to the adoption of privacy-enhancing technologies. A privacy-enhancing technology (PET) can be defined as: " a system of Information Communication Technology (ICT) measures protecting informational privacy by eliminating or minimizing personal data thereby preventing unnecessary or unwanted processing of personal data, without the loss of the functionality of the information system" [28].

PETs include, among others, data-anonymization methods, data-separation methods and authorization concepts that are decided upon by users and access controls. PETs allow individuals to make decisions about how their personal data can be processed and, most importantly, accommodate the principle of data minimization for personal information. Next, we present the background concepts and terminologies related to this thesis.

1. ***Goal modeling:*** One of the biggest challenges of privacy modeling is related to the adequate identification of privacy goals that do not negatively interfere with the functionality of mobile applications. To date, existing approaches have focused on goal oriented frameworks in the security domain, for the analysis and design of system software. These include UML, the Knowledge Acquisition in Automated Framework (KAOS) [30], the i* framework [31], Tropos and NFRF framework as explained in [32]. Our research focuses on privacy goal modeling to analyse privacy goals and concerns in mobile applications.

2. ***Privacy requirements modeling:*** Several privacy requirements methodologies have been employed in the literature, for example, LINDDUN [33] and Pris [34]. However, designing systems to meet privacy requirements is challenging and this is why privacy modeling is inevitable. Initiatives like the *The fair Information Practice Principles (FIPP)* privacy guidelines [35] are therefore important to ensure that privacy measures are met. FIPPs are used as a guideline to protect privacy especially how online companies collect and use personal data. Our approach however uses a privacy and functionality trade-off matrix to improve privacy design in mobile apps.

- Requirements modelling in Automotive industrial standards: Common practices in requirements engineering focuses on functional requirements, functional safety requirements and security requirements. In the automotive industrial settings, goals are defined using Automotive SPICE which is a status determination and risk assessment tool [36]. In addition to these methodologies, requirements engineering management is performed with the management tool: IBM Dynamic Object Oriented Requirements Management System (DOORS) Next generation. We employed IBM DOORS for requirements management in V2X telematics application design.
- Requirements modelling in eHealth apps: Google Play Store has guidelines on designing privacy policies for all apps that are to be published on the Google play website. We observed a variety of tools used to design requirements in current apps ranging from using mockups tools like mockplus [37] to using wireframes with tools like Balsamiq [38] for designing the requirements specification document. We applied mockplus to model requirements for eHealth applications. However, more research is required to design privacy requirements during the initial stages of eHealth app development.

3. ***Trade-off analysis methods:*** A trade off analysis explores situations in which factors are not attainable at the same time. This must be solved by balancing two opposite situations or qualities to get a best match. Several approaches have been proposed to model and analyse architectural and security requirements trade-offs. These include the Architecture trade-off Analysis (ATAM) [39], and the Security Verification and solution Design Trade-off analysis approach (SVDT) [40]. However, managing privacy trade-offs has not been extensively studied in existing literature. Our approach focuses on the design of a privacy trade-off at the initial stages of mobile applications development with a focus on app functionality, data collection and privacy policies and how these impact privacy preservation in modern applications.

2.1 Privacy-by-design concepts

This section presents three major privacy design concepts which are based on the paradigm of designing mobile application by implementing privacy-by-default. The term privacy-by-default in this thesis, relates to designing mobile applications according to *privacy-by-architecture* and *privacy-by-policy*. These privacy design concepts are crucial in implementing privacy and data protection at the initial stages of system development. They are part of the privacy and data protection law and emphasize that privacy should be considered upfront while designing mobile applications. Technical and organizational measures should be implemented in advance in the development life-cycle to ensure user privacy is protected. In particular this means that companies have to integrate privacy protection technologies through the whole system development cycle. Privacy has to be integrated by default in the seven phases of the development life-cycle which are; system planning, requirements analysis, system design, system development, testing, system deployment and system maintenance. Companies and organizations that collect, store, control and process data must be committed to establish a privacy aware culture through out the whole companies processes. This must be done for example, by training staff on privacy protection measures and techniques. The privacy-by default concept, therefore, plays a significant role in designing current mobile applications due to the massive data collection involved. To curb privacy concerns significant efforts have been made in research and development to improve these concepts [27] [41], [42]. This thesis, explores to which extent these concepts have been implemented in current mobile applications. In the following, we will elaborate in detail on these concepts to clearly understand the related background terminologies used. First, we outline some of the organizations and bodies that regulate privacy development.

Privacy regulation bodies and principles There exists a number of privacy regulation bodies and privacy principles that have been developed to assist in the implementation of privacy in both industry and legal frameworks. The most recent development that has attracted major attention is the introduction of the EU General Data Protection Regulation (GDPR) 2016/697, in May 2018. GDPR is discussed extensively in section

2.4. Several other principles also exist, these include Fair Information Practice Principles (FIPPs). FIPPs are a set of standards used to monitor the collection of personal data across different countries. They comprise eight major principles which guide companies on how they may use personal information in business transaction processes. Another privacy implementation body's principles worth mentioning are the Organization for Economic Cooperation and Development (OECD) privacy principles, which consist of privacy and data protection laws agreed upon by OECD member countries. Similar principles have been generated by the Asia-Pacific Economic Cooperation (APEC). APEC devised the APEC privacy framework to safeguard information privacy across its 21 member states. The protection of privacy has become a major concern, especially as companies have continued to collect users' data without their consent. The current increase in privacy breaches has, therefore, resulted in governments and legislature bodies, working in partnership with commercial bodies, intervening in order to protect user data. A set of key principles that have been adopted are the Generally Accepted Privacy Principles (GAPPs) [43]. GAPPs were developed by a privacy task force of the American Institute of Certified Public Accountants and the Canadian Institute of Chartered Accountants to address the challenges companies were facing through privacy breaches. GAPPs comprise 10 major principles which organizations can use to protect user privacy, especially personal identifiable information and protected health information.

These organizations play a major role in regulating privacy preservation on a general basis, we will now give an overview of the privacy-by-design concepts used in this thesis in detail, that is privacy-by-architecture and privacy-by-policy.

2.1.1 Privacy-by-architecture

Privacy-by-architecture approaches aim at protecting data by using techniques like cryptography, zero-knowledge proofs and pseudonyms, that prevent attackers from accessing data. A pseudonym is an identifier of a subject, in the setting of the sender or recipient, other than one of the subject's real name. A subject is pseudonymous if a pseudonym is used as an identifier instead of one of its real names. Pseudonyms are a set of short-term

identifiers that ensure privacy of the user by hiding the real identity of the sender. The privacy-by-architecture approach was initially introduced in [27]. Spiekermann and Cranor presented guidelines for building privacy-friendly systems using *privacy-by-architecture*. The method for addressing *privacy-by-architecture* minimizes the collection of personal data and emphasizes randomization, client-side data storage and processing. This involves using client-centric architectures which embed anonymous transactions that prevent adversaries from performing unauthorized access controls. Privacy-by-architecture employs anonymity based technologies which focus on protecting user identity. In practice, pseudonyms is used to protect Personally Identifiable Information (PII).

Data obfuscation technologies are further used to prevent linking PII to individuals. This may be done by creating ambiguous applications to prevent reverse engineering applications and linking information back to the source. Several research approaches have been proposed to design systems using privacy-by-architecture techniques especially pseudonyms. There are four major categories of pseudonym schemes, namely: (1) based on asymmetric cryptography, (2) based on identity-based cryptography, (3) based on group signatures, (4) based on symmetric cryptography. In the following we elaborate on some of the research where pseudonyms have been used to design applications based on the privacy-by-architecture concept in the vehicular communication domain.

Petit et al. [44] conduct a survey of pseudonym schemes in vehicular networks. The authors categorize pseudonym schemes based on the cryptographic means they employ. The four major categories are schemes based on asymmetric cryptography, schemes based on identity-based cryptography, schemes based on symmetric cryptography and group-signature schemes. They propose an abstract pseudonym life cycle, make a comparison of the various schemes and present their advantages and disadvantages. However, there is no detailed discussion of multi-domain privacy in this work.

Feiri et al. [45] outline problems specific to the real-world deployment of pseudonyms in vehicular networks. The authors find that using pseudonyms does not guarantee privacy, as vehicles are attached to short-term identities to provide accountability and non-repudiation. Although passengers expect privacy in the form of anonymity, pseudonym-change strategies

only offer unlinkability with small immobile adversaries. They stress that it might be necessary to find out whether the cost and complexity of using pseudonyms is worth the limited level of privacy achieved. This paper does not provide any evaluation concerning privacy in cross-domain settings of V2X communication.

Fonseca et al. [46] propose a multi-layer scheme which provides privacy at different layers and packet-forwarding schemes that employ packet-caching. Their scheme employs an extended location service which uses a secure routing protocol to provide message authentication based on digital signatures and asymmetric cryptography. They present four major solutions based on: (1) a cross-layer addressing concept (2) an extended location service (3) pseudonymity-enhanced packet forwarding and (4) link-layer callbacks. Their proof of concept of an OBU communication system observed that a delay in the packet forwarding-process during pseudonym changes can be reduced to enable unicast communications. Although this approach offers an enhanced packet-forwarding scheme based on pseudonym-caching, this may lead to an increase in the communication overhead. In addition, it employs a unicast wireless multi-hop connection between two nodes and geocast broadcasting of data packets, which are susceptible to geo-location tracking, thus endangering users' privacy.

Foerster et al. [47] present three privacy-preserving approaches, namely, the basic pseudonym scheme, the PUCA scheme and a credentials-based approach. The PUCA scheme implements full anonymity and uses anonymous credentials for privacy-friendly authentication with back-end providers when requesting pseudonym certificates. A question left unanswered is how privacy-preserving mechanisms used to protect back-end servers, especially for users data traversing multiple domains, are to be implemented.

Armknecht et al. [48] propose the use of PKI+ as a solution to enhancing cross-layer privacy. They observe that the use of PKI+ has an advantage due to the small data size resulting from the revocation of all its certificates, thus size of the communication overhead. Using PKI+ provides users with anonymity by applying OBUs that are linked to GIDs by an authority. Tracking is avoided by changing pseudonyms frequently, and since vehicles generate there own pseudonyms an adversary cannot link old pseudonyms

to new ones. Though this paper claims good cross-layer privacy results by employing PKI+, the paper does not provide an evaluation of the approach.

In [49], the Peer-to-Peer Anonymous Authentication (PPAA) scheme based on a group-signature scheme is proposed, in which a VANET is considered to be a P2P system. PPAA focuses on privacy and accountability; more notably, both clients and servers are peer users with privacy concerns. The study's major contribution is that peers are pseudonymous to one another but are anonymous and unlinkable to other peers in the network. This results in a high level of anonymity between peers, though a question that remains to be answered is how the revoking of misbehaving vehicles is to be solved.

Sun et al. [50] propose a pseudonymous authentication scheme (PASS), which reduces the revocation-cost and certificate-updating overhead by using hash chains to reduce the certificate revocation list (CRL). Their approach employs a proxy re-signature technique used for updating certificates. The authors disclose that in comparison to other related schemes, their approach has the smallest communication overhead and provides unlinkability between RSU and vehicles. However, it is not clear how the certificate authority (CA) is to manage issued certificates without having full control. The problem of having a single point of failure at the CA, which may be used to attack the system, still exists.

Weimerskirch [51] distinguishes privacy in two major categories: Privacy against 3rd party entities and privacy against authorities. Privacy against 3rd party entities involves the use of pseudonyms to ensure anonymity and unlinkability. Privacy against authorities involves sharing of power by V2X authorities (e.g. certificate authority and registration authority) to enable the recovery of privacy-sensitive information. The paper neither provides an implementation nor an evaluation of the solutions mentioned, merely outlining concerns over privacy deployment in V2X systems.

Widersheim et al. [52] present an exploratory study of an attacker using a Multiple Hypothesis Tracking (MHT) approach to vehicular networks. The results show that even with changing pseudonyms, an attacker is able to track vehicles and relate pseudonyms to specific drivers, even under noisy data, thereby raising questions about the effectiveness of pseudonymous schemes in V2X networks and the level of achievable privacy protection.

Lu et al. present. [53] a pseudonym-changing strategy (PCS), which aims to achieve a high level of location privacy in areas where a lot of vehicles gather, for example at intersections, in parking areas or at traffic lights termed social spots. They develop anonymity-set analytical models and employ game theory to investigate the privacy levels attained by using a PCS strategy. Though their evaluation of changing pseudonyms at a small social spot (e.g an intersection) and a large large social spot (e.g a parking area), they attempt to assess the problem of location privacy; however, there is lack of detail to solve the problem of privacy protection in mixed zones for both small and large social spots.

Qu et al. [54] present a review of VANETs and outline methods for providing vehicles with changing pseudonyms, which include preloading pseudonyms in tamper-proof devices (TPDs). The authors point out that these mechanisms are insufficient because vehicles need a large storage capacity to keep anonymous public keys. In addition, revoking malicious nodes uses a large storage space. The paper further notes that having pseudonyms using RSUs is insufficient if the limited wireless channel bandwidth is taken into consideration, because transmitting hundreds of certificates is difficult, especially with peak vehicle densities. The authors observe that using authentication to secure VANETS poses privacy risks to users and that many privacy protocols improve privacy at the cost of safety, as demonstrated by random silent periods. The paper conducts a review and provides no evaluation of the proposed methods for changing pseudonyms.

Feiri et al. [45] outline problems specific to the real-world deployment of pseudonyms in vehicular networks. The authors find that using pseudonyms does not guarantee privacy, as vehicles are attached to short-term identities to provide accountability and non-repudiation. Although passengers expect privacy in the form of anonymity, pseudonym-change strategies only offer unlinkability with small immobile adversaries. They stress that it might be necessary to find out whether the cost and complexity of using pseudonyms is worth the limited level of privacy achieved. This paper does not provide any evaluation concerning privacy in cross-domain settings of V2X communication.

Song et al. [55] explore challenges and solutions arising from using pseudonym

technology for privacy protection in e-services. Their paper examines two types of pseudonym-technology requirements, namely, privacy-related requirements and security-related requirements. Under privacy-related requirements fall pseudonymity, unlinkability and property-sharing resistance. Security-related requirements, on the other hand, include authentication, security of users' secret keys and the security of protocols. The paper further lists important pseudonym technologies, such as e-cash, e-ticket and e-voting, and describes privacy applications that have been implemented in these systems to provide anonymity and unlinkability. However, there is no evaluation of the privacy technologies applied.

Karragianis et al. [56] undertake a detailed analysis of vehicular networks, including their characteristics, standardization efforts, projects undertaken (in Europe, Japan and USA) and finally the challenges faced in vehicular communication environments. Among other details, the authors observe that one of the solutions of solving linkability between pseudonyms is to use silent periods and the creation of groups that prevent messages sent in one group being listened to by vehicles in another group. The second solution given is to use mixed zones, which share a secret key. If vehicles leave the mixed zone, keys are exchanged, thus protecting location privacy. This survey, however, does not take into consideration the issue of data privacy in multiple domains of vehicular communication components.

Analogous to the vehicular domain, current mobile applications gather a lot sensitive data and face numerous privacy challenges. There is growing concern about the immense amounts of private data being collected and processed; thus, in an attempt to reduce potential privacy risks, the key challenge is to safeguard personal data during data flows across mobile application processes. This is the major drive behind this thesis. Much as anonymisation techniques have been used in other domains like mobile networks and Intellectual property to protect privacy by privacy-by-architecture, data collection companies continue to collect PII without such measures being put in place. The major drive behind this phenomenon is the use of online target advertising which greatly compromises user privacy. Therefore, we advocate for mobile application designers to

consider privacy protection while designing mobile application that involve sensitive data traversing multiple sources and domains during system development. Subsequently, mobile application development has to be regulated by legal entities that ensure that they are built based on standardized solutions to close the gap of building applications without integrating privacy-by-architecture.

2.1.2 Privacy-by-policy

Privacy-by-policy as described in [27] focuses on the notice and choice principle. Notice is the presentation of terms used in a document which may be presented as a privacy document with terms of the agreement or a privacy policy. Choice on the other hand is the the action of accepting the terms and conditions offered in the privacy document or policy by either signing the privacy document or clicking on an agree button e.g in an privacy policy given to data subjects online. User privacy depends on how the policies governing a system are integrated and managed in an organization. The notice and choice principle involves notifying data subjects of how there data is going to be used. They have a choice to either consent or decline. This highly depends on the way the system is designed and managed by system operators to enable user control, access control to collected data. In practice, companies tend to employ privacy-by-policy ordinarily than privacy-by-architecture as system administrators often lack the knowledge to design privacy preserving technologies. Therefore, there is a gap in educating system managers on how privacy can be enhanced in mobile applications. The other major gap facing current mobile applications is the assumption that the giant data collecting corporations can be trusted to protect the data they collect. Likewise, data subjects assume that the documents they sign while providing PII e.g. privacy policies can be enforced by government legal entities. This is not always the case as we have seen a number of privacy breaches and misuse done by giant corporations that have not been penalized. Therefore, data subjects have to be enlightened on privacy protection methodologies so that they can choose which type of data is subjected to data collecting companies. Next, we will elaborate more on the implementation of the privacy-by-policy concepts that have been

initiated legally as an effort to protect user privacy.

Privacy-by-policy from a legal perspective: Privacy-by-policy is regulated by legal entities in order for the privacy documents and privacy policies to be effective. This thesis is based on the General Data Protection regulation (GDPR) and focuses mainly on the EU regulations. We have grouped the into two categories : legislature based on *Notice principles* and legislature based on Choice principles.

EU data protection regulation from the perspective of Notice principles: The EU data protection laws stipulate that data controllers have to inform data subjects about the data they collect and this is legally provided under article 14 of the General data protection regulation (GDPR) [57]. Below is an excerpt of article 14 which explains the legal measures to be taken before processing data.

1. Where personal data have not been obtained from the data subject, the controller shall provide the data subject with the following information:

   (a) the identity and the contact details of the controller and, where applicable, of the controller's representative;

   (b) the contact details of the data protection officer, where applicable;

   (c) the purposes of the processing for which the personal data are intended as well as the legal basis for the processing;

   (d) the categories of personal data concerned;

   (e) the recipients or categories of recipients of the personal data, if any;

   (f) where applicable, that the controller intends to transfer personal data to a recipient in a third country or international organization and the existence or absence of an adequacy decision by the Commission, or in the case of transfers referred to in Article 46 or 47, or the second sub paragraph of Article 49(1), reference to the appropriate or suitable safeguards and the means to obtain a copy of them or where they have been made available.

2. In addition to the information referred to in paragraph 1, the controller shall

provide the data subject with the following information necessary to ensure fair and transparent processing in respect of the data subject:

(a) the period for which the personal data will be stored, or if that is not possible, the criteria used to determine that period;

(b) where the processing is based on point (f) of Article 6(1), the legitimate interests pursued by the controller or by a third party;

(c) the existence of the right to request from the controller access to and rectification or erasure of personal data or restriction of processing concerning the data subject and to object to processing as well as the right to data portability;

(d) where processing is based on point (a) of Article 6(1) or point (a) of Article 9(2), the existence of the right to withdraw consent at any time, without affecting the lawfulness of processing based on consent before its withdrawal;

(e) the right to lodge a complaint with a supervisory authority;

(f) from which source the personal data originate, and if applicable, whether it came from publicly accessible sources;

(g) the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.

3. The controller shall provide the information referred to in paragraphs 1 and 2:

(a) within a reasonable period after obtaining the personal data, but at the latest within one month, having regard to the specific circumstances in which the personal data are processed;

(b) if the personal data are to be used for communication with the data subject, at the latest at the time of the first communication to that data subject; or

(c) if a disclosure to another recipient is envisaged, at the latest when the personal data are first disclosed.

4. Where the controller intends to further process the personal data for a purpose other than that for which the personal data were obtained, the controller shall provide the data subject prior to that further processing with information on that other purpose and with any relevant further information as referred to in paragraph 2.

5. Paragraphs 1 to 4 shall not apply where and insofar as:

   (a) the data subject already has the information;

   (b) the provision of such information proves impossible or would involve a disproportionate effort, in particular for processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, subject to the conditions and safeguards referred to in Article 89(1) or in so far as the obligation referred to in paragraph 1 of this Article is likely to render impossible or seriously impair the achievement of the objectives of that processing. In such cases the controller shall take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available;

   (c) obtaining or disclosure is expressly laid down by Union or Member State law to which the controller is subject and which provides appropriate measures to protect the data subject's legitimate interests; or

   (d) where the personal data must remain confidential subject to an obligation of professional secrecy regulated by Union or Member State law, including a statutory obligation of secrecy.

EU data protection regulation from the perspective of Choice principles: Lawful data processing must be done subsequent to data subjects consent. This means that data recipients have to explicitly inform data subjects of the intentions of how they are to process data. Consent is referenced in article 7 of the General Data Protection regulation as follows;

1. Where processing is based on consent, the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data.

2. If the data subject's consent is given in the context of a written declaration which also concerns other matters, the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language. Any part of such a declaration which constitutes an infringement of this Regulation shall not be binding.

3. The data subject shall have the right to withdraw his or her consent at any time. The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal. Prior to giving consent, the data subject shall be informed thereof. It shall be as easy to withdraw as to give consent.

4. When assessing whether consent is freely given, utmost account shall be taken of whether, interalia, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract.

At the time of writing this thesis a year after the GDPR was enacted, large mobile application corporations like Google and Facebook have not fully integrated its regulations. Subsequently, CNIL, the data protection regulator in France have issued a €50 million penalty to Google for not complying to the GDPR regulations [58]. With this limitation of existing privacy protecting approaches in current mobile applications, we call for stringent enforcement of GDPR regulations and other data privacy regulation laws in current and future mobile applications.

Hybrid approach: In [27], guidelines for building privacy-friendly systems using two approaches: *privacy-by-architecture* and *privacy-by-policy* were presented. The method for addressing *privacy-by-architecture* minimizes the collection of personal data and emphasizes randomization, client-side data storage and processing. *The privacy-by-policy* method, on the other hand, focuses on the implementation of notice-and-choice principles for fair information practices. The authors further, point out that, according to current IT architectures, there are three technical domains for building privacy-friendly technologies and information systems: *user sphere*, *recipient sphere* and *joint sphere*. The *user sphere*

encloses a user device, it should be fully controllable by the user who owns it and data should not flow without the user being able to control it. The *recipient sphere* is a company-centric sphere of data control that involves back-end infrastructure and data-sharing networks. The *joint sphere* includes companies that host people's data and provide additional services.

Enhancing privacy across different domains can be analyzed via the following key questions:

- How is individual privacy maintained when data are crossing many boundaries (data in motion)[1], that is, from the user-sphere domain to joint and recipient sphere domains?
- How do we ensure individual privacy for data at rest [2] and avoid secondary and tertiary(n-ary) further processing of data in the joint and recipient spheres?
- How do we ensure that the trade-off between application functionality(effectivity) and privacy is minimized to improve customers' confidence in emerging mobile applications?

We take into account the approach presented in [27] to design privacy-preserving systems using *privacy-by-architecture* and *privacy-by-policy* which play a crucial role in building mobile applications, especially in the early stages of development. In general, a hybrid of the two concepts is required for building robust privacy aware mobile applications [60].

Therefore, there is a need for a clear-cut method of implementing a hybrid of *privacy-by-architecture* and *privacy-by-policy* techniques in mobile applications to ensure that policies are put in control over their personal data. We refer to user-privacy in mobile applications as the protection of information, such as location data, name, driving route and user identity, which can be used by an adversary for illegal tracing and user-profiling.

In order to achieve this goal, we need to do the following:

[1]Data in Motion are data that are traversing the network, or temporarily residing in a computer memory to be read, updated or forwarded to another data-processing service [59].

[2]Data at Rest refer to 'inactive data' physically stored in databases, data warehouses, spreadsheets, archives, tapes, offsite backups or on mobile devices [59].

1. Specify requirements for implementing mobile application design processes using *privacy-by-architecture* and *privacy-by-policy*.

2. Identify clearly what the *privacy-by-architecture* and *privacy-by-policy* principles mean in the mobile application domain.

3. Specify the *privacy-by-architecture* and *privacy-by-policy* best practices applicable in mobile applications and make them accessible to policymakers and design engineers.

Next, we will elaborate on privacy challenges involved in designing mobile applications.

2.2 Privacy-by-design challenges

Data-management systems typically perform data transfers, data storage and data processing. With this in mind it is important to understand how privacy breaches can occur and how to counter them. Spiekermann and Cranor [27] analysed privacy requirements on the basis of privacy sensitive processes, user perceptions and concerns. Privacy concerns can be subdivided into three major spheres. First, users are concerned about data being collected and processed in the user sphere or their personal systems without user awareness. In particular, privacy concerns arise over unauthorized collection, unauthorized execution, exposure and unwanted inflows and outflows of data. Second, after a transfer of data from the user sphere, they are hosted by companies in the joint sphere. Here, users do not have any control over their data and, therefore, they are concerned about their data being exposed, unauthorized secondary use and improper access by external third parties. Third, data storage in back-end infrastructure systems is a major concern as data are stored without users' consent and so users need to trust the recipients in the remote sphere. Therefore, steps must be taken to ensure that internal and external unauthorized uses of data, improper access to data by third parties, and errors in data are avoided.

Spiekerman and Cranors' suggestion to design privacy-preserving systems using *privacy-by-architecture* and *privacy-by-policy* has similarities with our research; however, prior work has documented that privacy-by-design is a vague concept [61] which leaves many open issues regarding the application of its methods and the translation of its principles

to actual designs of system engineering. For example, each company or commercial party can decide individually how privacy-by-design is implemented. Much as this approach is innovative, it does come with certain disadvantages, e.g. users, such as policymakers, local authorities etc, have to understand how each company implements privacy-by-design.

Furthermore, existing privacy approaches, such as privacy-by-design or PETs, are, however, still challenging when directly transferred to mobile communication systems due to the way in which these systems are built and implemented. The distributed nature of mobile applications makes existing privacy approaches that are implemented for different stakeholders or entities very difficult to assess because of random mobility and frequent disconnection.

To enforce personal data privacy, in particular Personally Identifiable Information (PII) across multiple parties, Pearson et al. propose sticky policies, which may also be considered a privacy-by-design approach. Sticky policies (disclosure policies) can be used to allow the selective disclosure of any aggregation or combination of confidential information [15]. Protecting PII is vital, because it contains data that can be traced to a particular individual. Such data are sensitive and may contain: passwords, email addresses, social security numbers, personal identification numbers, medical data or financial data. The authors state that much as substantial research has been conducted to provide mechanisms for online privacy management, major issues remain, such as how to give more control to end users and how to gather and manage end-users' content. In this regard, an approach based on sticky policies was determined. Sticky policies are conditions and constraints attached to data that describe how data are to be used. Sticky policies help in controlling how data are to be accessed and used, or how data are accompanied through an entire distributed system [16]. The use of sticky policies is a good idea for protecting user data; however, previous research has mainly focused on encryption techniques that can finally be traced back to the user. Thus, sticky policies are not enough to effectively handle secondary and tertiary(n-ary) scenarios because there still exist privacy risks related to the abuse of information by parties with legitimate access to back-end servers and databases. There also remains a need for an effective method of protecting user data during information flows

across organizational boundaries and cross-domain boundaries in the system development process. Cross-domain privacy is the process of protecting information transmitted between different domains which have different levels of mutual trust. This refers to user data that flows from a mobile device or application to other organizational boundaries and entities in the mobile communication architecture. Mobile communication systems are composed of largely independent subsystems with multiple domain boundaries, making it hard for user privacy to be maintained across different domains. There still exists a trade-off between privacy and application functionality(effectivity) when developing mobile communication systems. Much as most of us are aware that functionality (effectivity) and privacy are desirable, we understand that enjoying one will lead to giving up the other. This trade-off has been part of our lives and our choices. However, corporations are increasingly building user profiles without users' knowledge and thus users are gradually losing the option of choice. Similarly there is a trade-off between privacy and security. As users reflect a heightened concern for security, the *nothing-to-hide* argument often declared by users necessitates a higher demand for security than privacy. This means users are willing to give out their personal information as long as they have nothing to hide or, to put it differently, if users feel they only engage in legal activity, then they do not have to worry about surveillance or data-mining. The *nothing-to-hide* argument has been extensively discussed by Solove [62]. However, there is little research on the impact of the nothing-to-hide phenomenon or the trade-off between functionality and privacy in current mobile-application systems.

Our research will also target this problem area. In particular, we argue that user data collection, user data processing and user data dissemination increase the vulnerability to potential abuse of user information as users do not have a say in how their data are processed. The key challenge in our research is the privacy of user data during data flows across the entire mobile application architecture. The hyper-connectivity and constant distribution of messages between mobile applications involves multiple privacy risks. For example, the manipulation of data on a mobile device may cause problems for applications running on other connected devices, resulting in privacy infringements. Also, mobile

applications require the collection of massive amounts of data, e.g. past and present location data. This data is aggregated and sent to back-end systems, such as data control centers, for further processing. This type of data collection poses privacy challenges; for instance, data crossing domain boundaries can be intercepted when being sent from a mobile application to its destination. Similarly, data stored in back-end databases can be manipulated. The large volumes of data collected make them more appealing to attackers. Mobile applications involve multiple actors, meaning that secondary/tertiary leakage or abuse of personal information is highly possible because the user has no control in terms of how their data are used at secondary and tertiary levels. Therefore, there is an urgent need to design privacy preserving mobile applications that provide cross-domain data security and privacy. Thus, preventing unauthorized secondary/tertiary usage of data for the successful deployment of mobile applications is key.

The aforementioned privacy vulnerabilities indicate that personal information in mobile applications is not as secure or inaccessible as users may assume. There exist implicit possibilities for user privacy breaches, such as tracking users' movements and the generation of user profiles through privacy and security leakages. In addition, it is not only security leakages that pose privacy threats, the systems accessing users' data can also enable profiling, and this can be disastrous if these data are leaked to less ethical organizations and adversaries. The following section provides privacy engineering methodologies that are used to design some of the current privacy preserving applications.

2.3 Privacy design strategies

Hoepeman introduces eight privacy design strategies which are derived from Spiekermanns and Cranors' framework for privacy-friendly system design. We will now discuss the eight strategies from a mobile application design perspective. in relation to designing mobile applications.

- **Minimise:** The first strategy is to minimise and it states that: "*The amount of personal information that is processed should be minimal.*" This strategy is in alignment

with our approach to designing mobile applications, especially where companies are reported to continuously collect data and personal identifiable information without consent from data subjects. We emphasize minimal data collection by modern apps, especially so if users can be identified and monitored when using mobile apps.

- **Hide:** The hide strategy states that: " *Any personal information that is processed should be hidden from plain view.*" In our view, in most mobile applications data are transferred to back-end servers without any form of anonymity or data-hiding techniques. This has led to the possibility of personal data being compromised, although initial authentication, e.g. by using passwords, has been done on the client side. Also, data sets in databases are not anonymised. Therefore, this strategy should be implemented to reduce privacy data breaches that occur in both back-end servers and mobile application application databases.

- **Separate:** The separate strategy states that: "*The processing of personal information should be done in a distributed fashion whenever possible.*" Privacy is best implemented with decentralized systems which do not allow the generation of user profiles. Distributed mobile applications should therefore be in a position to run on different database systems so that the tracking of data subjects is avoided.

- **Aggregate:** Aggregation is an important strategy in the design of mobile applications. The aggregate strategy states that: "*Personal information should be processed at the highest level of aggregation and with the least possible detail such that it is (still) useful*". Once personal information is aggregated it makes it harder for adversaries to identify and monitor their subjects. A good example of such an implementation is the changes that STRAVA put in place after a heat-map data dump revealed location information about its users. The map zoom level was reduced so that maps were not able to reveal users' precise locations without further authentication. Therefore, aggregating data to protect user privacy plays a crucial role in implementing privacy in mobile applications.

- **Inform:** This strategy is mentioned in most well-known privacy principles. It states

that: "*Data subjects should be adequately informed whenever personal information is processed.*" However, the industry has not been keen on informing data subjects about the personal information that is collected. In most cases, regarding what is seen in privacy policies, data subjects are required to pay a fee in order to get details about their personal data collected by companies. Information about personal data collected should therefore be accessible without paying any service fees so as to improve trust and openness in mobile applications

- **Control:** The control strategy states that " *Data subjects should have agency over the processing of their personal information.*" Data subjects should be able to check how their personal data are used. Regulation has to be imposed on the collecting companies so that checks and measures are put in place to ensure that privacy principles are adhered to.

- **Enforce:** The enforce strategy states that: "*A privacy policy compatible with legal requirements should be in place and be enforced.*" This is in sync with what we advocate in this research. Our focus is on how privacy policies are implemented. In current mobile apps, privacy policies are not in sync with the data collected by mobile-application companies. Therefore, regulations that require the enforcement of privacy policies have to be enacted, not only in privacy policy statements but also in how collected data are used and processed.

- **Demonstrate:** The demonstrate strategy states that: "*Companies must be able to demonstrate compliance with their privacy policy and any applicable legal requirements.*" Compliance is a crucial step in privacy implementation, especially from a legal perspective. Companies should be able to prove that their systems are privacy compliant and adhere to privacy principles and guidelines. Companies may be compliant through certification, which is one way in which users may determine which companies they can trust to handle their personal data in a privacy-preserving manner.

2.4 Privacy design and legal aspects

The EU general data protection regulation (GDPR) [57] was effected on 25 May 2018. It contains 99 articles and 173 recitals. GDPR's major aim is to protect user data and privacy in European Union member states and during cross-border data transfers to third countries. After taking effect many companies have redesigned their privacy statements to meet the privacy demands proposed by GDPR. In the following we highlight some of the articles that are given major emphasis in this thesis.

- **Article 16: Right to rectification:**

  "*The data subject shall have the right to obtain from the controller without undue delay the rectification of inaccurate personal data concerning him or her..*"

  This article refers to the ability of the data subject to rectify inaccurate data that have been collected by data controllers without delay. This, however, has recently been massively abused in mobile applications, as seen in the rising cases of "fake news". The fake news phenomenon has proven very hard to rectify and critics say that it has turned into a major security risk. Spreading false news for personal gain has drastically increased, especially through social media platforms, and has, unfortunately, proved very hard to correct. Therefore, the right to rectification plays a major role in the design of modern mobile applications and should be integrated into the application development process by default.

- **Article 17: Right to erasure(right to be forgotten):**

  "*The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay.*"

  The right to erasure or to be forgotten plays an important role in the design of mobile applications. This involves the deletion of data that are considered old and outdated according to the duration for which they have been published. Examples of such data may include convictions of data subjects, unwanted information about individuals published through company URLs that could be used against them

when buying say insurance, taking out a mortgage or searching for jobs. There have been a couple of cases where data subjects have complained of their data online being erased by data controllers, such as Google, as in the case of Google Spain SL v Agencia Espanola de Proteccion de Datos. As a result, measures have been implemented by data-collection firms, such as Facebook and Google, to remove URLs in a bid to implement the right to erasure. In this thesis, we emphasize the right to be forgotten or the the right to erasure, especially in mobile communication applications where the data subject does not have full deletion control over their data held by data-collecting companies. We observed that most mobile applications to date had not put this into effect. Therefore, more attention should be paid to implementing data-erasure policies and techniques in the initial stages of application development.

- **Article 20: Right to data portability:**

  "*The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format...*"

  Data portability involves data subjects requesting information about the data being collected about them. This article plays an important role and ensures that data subjects are well informed about the data they submit to various data-collection companies. If needs be, they are able to delete them or request their deletion as envisaged in article 17. In this thesis we observe that most mobile applications still lack a mechanism to implement data-portability procedures. Some companies have explicitly included clauses about how to request information about the data collected from data subjects in their privacy policies. However, this always comes with a fee payable to the company. This means that data subjects are reluctant to request information about data being collected, especially if it requires paying a fee. Therefore, better measures have to be put in place to ensure that data-collection companies implement the right to data portability without charging data subjects.

- **Article 21: Right to Object:**

  "*The data subject shall have the right to object, on grounds relating to his or her particular situation, at any time to processing of personal data concerning him or her...*"

  Data subjects have a right to stop data controllers processing their data. This means that companies have to accept data subjects' demands and stop processing without any delay. However, data controllers do not have to delete data they have collected in the process. So this means that although companies can stop the data processing for, say, marketing purposes, they still retain the data. This is common in mobile applications, especially when companies retain data for research purposes. These data can easily be sold to third-party service providers who can in turn process them further. While the data have to be processed in a more secure way, using data minimisation principles and pseudonymisation techniques, many data breaches still occur, which thus calls for immediate attention in this field of research. In this research, we call for the deletion of all data collected once data subjects have requested a halt to processing them.

- **Article 22: Automated individual decision-making, including profiling:**

  "*The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.*"

  This article restricts data controllers from obtaining personal information for profiling using automated processing. This means that companies are not allowed to make decisions or collect any information that is used to classify users based on, for example, personal facts, health data, interests, lifestyle or where they live automatically, without any human intervention. Automated decision-making systems are on the increase, especially in social media platforms and search engines which use machine-learning and data-science algorithms to identify user preferences for use in targeted advertising. Article 22 therefore aims to protect data subjects against such automated decision-making processes. It requires data controllers to inform

data subjects of any automated decision-making done while collecting data.

- **Article 25: Data protection by design and by default:**

  "*..The controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed.*"

  This article is the main focus of this thesis. We argue that controllers should implement privacy-protection measures in the initial stages of development, and especially that privacy should be implemented by default and not as a response measure after privacy breaches have taken place. In this way data subjects are protected from the beginning of the data-collection procedure.

- **Article 32: Security of processing:**

  "*...the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk, including inter alia as appropriate.*"

  Data controllers and companies that process personal data must put in place technical security mechanisms that ensure that the security of user data is maintained. This may be done using policies and data-protection processes that ensure that risks are addressed and controls are put in place to protect users' data. In our research we mainly focus on data collected in mobile applications. Different apps should include mechanisms that protect user data not only at the app level but also during data transfer and in back end servers that store personal data. Risks should be minimised by introducing security controls that mitigate attacks. This can be done using pseudonyms or end-to-end security encryption, e.g. as implemented in mobile applications such as WhatsApp. In this way, data-transfer processing is secured against potential attackers especially for data in motion. Our research emphasizes the implementation of security controls that regularly assess if security measures are applied. To date, privacy and security techniques are not regularly used to enforce security in mobile apps.

- **Article 34: Communication of a personal data breach to the data subject:**

  "*When the personal data breach is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall communicate the personal data breach to the data subject without undue delay.*"

  This article refers to data collectors notifying data subjects of any data breaches that occur. In the event of any data leak, companies have to inform individuals about the leak and the risks involved. This can be communicated via email, as in the case of Yahoo! data breaches where hackers managed to compromise user accounts in 2013. The data leak exposed the email addresses, passwords, dates of birth, security questions, answers to security questions and hashed passwords of all 3 billion users. Such data breaches must be communicated to the users in order for data subjects to be able to change their credentials so that hackers are prevented from using them. As a result, Yahoo! took action and notified all users who were affected to change their passwords and security questions. This was also communicated on its website so that users could change their credentials. This article has therefore been well implemented by giant data controllers, but little has been done in the area of mobile applications to apply it. Our research therefore maintains that mobile application companies should inform data subjects whenever data breaches occur.

  In the following, we elaborate on some of the key privacy frameworks, described in literature, designed to enforce the aforementioned privacy principles.

2.5 Privacy-by-design frameworks

A number of privacy frameworks are described in the literature to aid developers implement privacy principles in system development. Among the well-established privacy frameworks in the literature is Ann Cavoukian's privacy by design (PbD) framework developed in the mid-nineties [63]. PbD has the ability to integrate privacy in all stages of development, ranging from application systems and network infrastructures to the business application

level. The general idea behind PbD is to integrate privacy by default, even when the user has not asked for any such implementation or does not intend to use it.

1. **The Privacy-by-design framework**

   The PbD framework is based on The fair Information Practice Principles (FIPP) principles described as follows:

   - *Proactive not reactive:* Privacy has to be implemented initially in the system before any privacy breaches occur, and not after attacks have been identified. Any privacy-invasive events have to be prevented before the system is in operation, not after it has been attacked.
   - *Privacy as the default setting:* Personal data in applications have to be protected by default without any need to change the settings or foreign action at both the business level and the networking level.
   - *Privacy embedded into design:* Privacy should be embedded in application design and not be an add-on, it should reside in the software and functional components of the application
   - *Full functionality (positive-sum, zero-sum) :* Privacy and security are both important when developing applications; therefore, both should be achieved without compromising any design goals of the system.
   - *Maintain end-to-end security:* Data should be secure in the whole development cycle, security should be maintained when data are in use and after they have been used and discarded.
   - *Visibility and transparency* Stakeholders should be aware of any business applications and technologies put in place to implement privacy. Privacy objectives should be met and it should be verified if they meet the required privacy principles.
   - *Have respect for user privacy and keep it user-centric :* Data subjects should be aware of how their data are processed by notifying them of any failures and offering them better privacy-protected solutions.

2. **OECD Privacy Framework**

The OECD privacy framework's main objective is to protect personal data and data flows across the borders of OECD member states and to enact laws that govern cross-border transfers of personal data. It comprises the following main principles:

Collection Limitation Principle: This principle aims to reduce the quantity of user-specific data that are collected by institutions, organisations and companies. The second part of this principle requires that users are aware of data being collected about them or are at least informed about data collected so that they may accept the conditions under which data are stored.

Data Quality Principle: At the heart of this principle is the notion that data acquired should be correct and constantly revised to ensure that they are accurate, especially for the aim fow which they are collected and intended.

Purpose Specification Principle: The main objective behind this principle is to state the reason why user data are being acquired. This has to be done before users hand over their personal data so that they can opt in or out in case they find that the reason for which data are being collected is not acceptable. For example, We have regularly seen many websites that are designed to collect personal addresses for sale, but without users being aware of the purpose of data collection.

Use Limitation Principle: This principle works in conjunction with the purpose specification principle and states that collected personal data should not be utilised other than for the purpose for which they were collected. The second part of this principle states that data should only be used only in cases where the data owner consents or where the law permits.

Openness Principle: This principle implies that how data are processed should be open, and principles and policies have to be put in place to monitor how data are processed. The principle also states that the people who process personal data should be known and accountable for the way data are processed and put means in place to protect collected data.

Security safeguards Principle: The security safeguard principle states that collected data should be protected by security mechanisms from attacks by malicious intruders. In addition, data should not be altered or misused by organizations that collect them.

3. **A Framework for Privacy-friendly System Design**

   Spiekermann and Cranor [27] produced guidelines for building privacy-friendly systems using two major approaches: Privacy-by-architecture and privacy-by-policy. *Privacy-by-architecture:*

   The method for addressing *privacy-by-architecture* minimizes the collection of personal data and puts emphasis on randomization, client-side data storage and processing. The method of *privacy-by-policy*, on the other hand, focuses on the implementation of notice and choice principles for fair information practices.

   We take into account the approach of designing privacy-preserving systems using *privacy-by-architecture* where *privacy-by-policy* plays a crucial role, especially in the early stages of development. Therefore, there is a need for a clear-cut method of implementing a hybrid of *privacy-by-architecture* and *privacy-by-policy* techniques to ensure that policies are put in place and users have control over their personal data. We refer to user privacy as the protection of personal identifiable information, such as name, home address, location data, email address, driving route and vehicle identity, which can be used by an adversary or third parties for illegal tracing and user-profiling. Next, we elaborate on some of the challenges of integrating privacy-by-design in mobile applications using the solutions mentioned above.

2.6 Privacy engineering methodologies

2.6.1 LINDDUN privacy threat modeling

A number of privacy methodologies have been proposed in the literature. LINDDUN privacy-threat modelling is one of the key methodologies used during our research to aid in generating privacy requirements. LINDDUN aims to identify privacy threats using

data-flow diagrams (DFD) of the system being investigated. They comprise six key steps derived from the STRIDE model [64].

- **Define data-flow diagrams(DFD:)**

  This is the initial step of the LINDDUN framework, it involves defining data-flow diagrams for the system under design. This process includes defining the system at a high level and distinguishing the elements in the system and how data flow in these elements, which are connected to each other.

- **Map privacy threats to DFD elements:**

  The next step entails mapping the DFD to privacy threats which are derived from privacy properties: linkability, identifiability, non-repudiation, detectability, information disclosure, content unawareness, policy and consent noncompliance.

- **Identify threat scenarios:**

  The third step of the LINDDUN framework is a core execution step, which involves using threat-tree patterns which can be used to attack a system's privacy. Again, these are based on the privacy properties listed above and are meant to protect the system from any privacy attacks.

- **Prioritize threats:**

  In this step, the most significant privacy threats are considered; these may be based on an assessment made by system and privacy engineers concerning how grave the impact of these threats might be. Risk assessments aid in determining which threats should be considered or not.

- **Elicit Privacy requirements:**

  Based on the threats identified in the threat prioritisation step, privacy requirements are generated using threat trees and misuse cases. Requirements are generated by determining the causes of threats and how these threats can be mitigated by making fine-grained changes.

- **Select corresponding privacy-enhancing technologies(PETS):**

The final step in this process is to decide on privacy-enhancing technologies and solutions to respond to threats. Which technologies are to be used in order to prevent privacy attacks hinges on the system under development. Privacy-enhancing solutions may further be divided into technical solutions which may be applied to the system or legal structures generated by government bodies which could be used to implement requirements.

2.6.2 Privacy enhancing technologies

Privacy enhancing technologies (PET) can be defined as a system of ICT measures protecting informational privacy by eliminating or minimizing personal data thereby preventing unnecessary or unwanted processing of personal data without the loss of the functionality of the information system [65]. PETS involve the use of data anonymization methods , data separation methods, authorization concepts that are determined by user access controls. PETS enable user empowerment on how individual personal data can be processed and most importantly allow for the principle of data minimization of personal information.

This involves the use of encryption technologies e.g, pseudonyms, digital signatures, blind signatures to protect user personal Identifiable information (PII). A number of PETS have been proposed in the literature which include digital signature algorithms e.g the Enhanced privacy ID (EPID), obfuscation methods and communication anonymizers. According to [65]. PETS can be classified depending on the capabilities and functions they offer. These include informed consent, data minimization , data tracking, anonymity and control. These capabilities play a major role in choosing which PET best suits the privacy challenge one is facing. We will now elaborate on these capabilities in relation to privacy design in mobile application focused on in this thesis.

- *Information Consent:* This thesis focuses on informed consent in the form of proper privacy policy design and implementation in mobile applications. Mobile application designers should inform users of the data stored, its use and how its shared before users consent to privacy policies given. Next, Data mimimization is key to the

reduction of the data being proceesed by mobile applications. Applications should process only the data required for system functionality and required services. Several attempts have been developed in this domain which include search engines that only collect infromation required for use to minimize user profiling e.g by storing web searches, IP iddresses, search histoty. Such engines aid the user by protecting PII and thus minimize data collection. These include Start page[66] , Duckduckgo[67], Disconnect Search [68], and Oscobo[69].

2.7 Privacy metrics

Privacy metrics in the literature can be defined as measures that determine the level of privacy in systems. Privacy metrics are used to determine which method of privacy implementation is best to use when integrating privacy into software applications. There has been a lot of research in the literature that aids in determining what level of privacy can be attained using privacy-enhancing technologies [70] [71]. In the following we describe some of the major privacy metrics that are reported in the literature. This thesis is not based on these metrics but we mention them as background concepts related to this thesis.

- **k-anonymity.** The k-anonymity privacy metric was introduced by Samaratti and Sweeney in 1998. k-anonymity uses tuples of attributes to uniquely identify individuals using quasi-identifiers [72]. The general idea behind it is to protect the identity of individuals within an anonymity set. That is, it minimises the possibility of re-identification of attributes in a data set by suppressing them or generalizing them so that they cannot be leaked to other data sets. The k-anonymity concept is commonly used to determine a data set's level of privacy. It has been recently used in combination with other cryptographic hashing techniques to determine the level of privacy of given passwords [73].

- **l-diversity.** l-diversity is an improvement over k-anonymity and was introduced by Machanavajjhala in 2007 [74]. He states that:

"An equivalent class is said to have l-diversity if there are at least l well-presented values for the sensitive attribute. A table is said to have l-diversity if every equivalence class of the tables l-diversity."

The l-diversity concept generates stronger levels of privacy compared to the k-anonymity approach as it provides a greater distribution of attributes within a data set Unlike the k-anonymity concept, the l-diversity concept improves anonymity in that attributes are not disclosed. It employs a better pruning mechanism compared to the k-anonymity approach through its l-diversity algorithm. However, critics of this approach maintain that l-diversity contains a lot of redundant attributes. It is also prone to privacy attacks, e.g skewness and similarity attacks [75].

- **e-differential privacy.** A brief definition of e-differential privacy from Machanavajjhala states that: "An algorithm satisfies e-differential privacy if its output on a database of individuals is statistically indistinguishable (measured by parameter e) from the output of the algorithm if any one individual had opted out of the database." e-differencial privacy aims to hide information about attributes in a database by using random values embedded in the results of applied queries. This approach improves the level of privacy such that query results are ambiguous and hard to reconstruct because of the random attributes [76].

- **Privacy metrics as presented by J. Bonneau and S. Preibusch.** Bonneau and Preibusch investigated online social networks to determine privacy practices and the implementation of privacy policies [77]. To evaluate the data collected based on 45 Online social networks, their research used imputed privacy metrics based on the amount of data collected, the privacy controls used and the design of the privacy policies. These metrics aimed at computing scores to determine privacy implementation in online social networks. The scores employed for data analysis are privacy scores and functionality scores. Privacy scores are computed using a data collection score, a privacy control score and a privacy policy score. Functionality scores are scores that represent features that are not required for system functionality in an online social network site. These scores are the basis of

the analysis of the data collected in our research and they play a significant role in the privacy preservation methodology used in our research.

Privacy Mobile Application framework Although various frameworks concerning privacy have been proposed in recent works e.g the above mentioned privacy frameworks; Privacy-by-Design (PbD) framework, OECD Privacy framework and the LINDDUN privacy framework, supporting frameworks for mobile applications are lacking. Extensive research regarding frameworks in the Internet of Things (IOT) domains have further been proposed in [78] and for detecting privacy violations in Android application code in [79], however, frameworks of incorporating privacy properties such as anonymity, unlinkability, unobservability, Pseudonymity, and confidentiality in the software development lifecycle of current mobile applications are missing. Furthermore, the integration of privacy protection and privacy policies through legislation bodies, to align with the fast development of current mobile applications has been not effective enough to prevent privacy abuse. Currently, there are no standards on how privacy is implemented in mobile apps making way for tech giants to prioritize their commercial and financial needs while developing applications. Although efforts have been made to integrate Privacy-by-Design in other domains like IOT systems [80] and in the medical domain [81], little research exists in the mobile app domain. Therefore the needs to be a standard method of protecting private data in mobile applications.

2.8 Summary

This section has presented the background and work related to this thesis. We focus on work presented in the literature that defines what privacy is and the role that privacy plays in protecting information about individuals. Our explicit interest in protecting user privacy has been discussed to give a clearer understanding to readers who do not know that they have a right not to have their personal data misused by companies. Companies should therefore, inform users about data being collected as envisaged by the GDPR data protection guidelines; furthermore, they should put in place structures that are secure to prevent any data breaches. For this research our intention is to illustrate

the approaches and mechanisms that are used to protect user privacy. These include the use of privacy-enhancing technologies , privacy-by-design which is a combination of privacy-by-architecture and privacy-by-policy principles and privacy-by-design frameworks. Companies should also abide by the rules and regulations that are imposed by regulatory bodies, such as the OECD and GDPR by implementing appropriate frameworks, e.g the OECD privacy framework and the framework for privacy-friendly system design.

We present some of the challenges faced when using privacy-by-design principles and privacy implementation in Telematics mobile applications by reporting on the survey of current telematics mobile application. This section further gives a short overview of the privacy metrics described in the literature.

The other limitations observed in the course of this study is that the concepts of privacy-by-architecture and privacy-by-policy are all applied and implemented individually in an attempt to improve user privacy. However, a hybrid approach is more effective where system engineers work with lawyers and HCI engineers to develop systems that are privacy friendly. Instead of, software engineers developing privacy-by-architecture solutions, privacy-by-policy strategies being formulated by lawyers and HCI engineers developing user interaction software there must be a way of all these solutions being consolidated together to provide a streamlined way of designing privacy preserving mobile applications. Personally Identifiable Information (PII) which includes sensitive information e.g Social Security numbers, National Insurance Numbers, Mailing addresses, Financial data, medical records and non sensitive information e.g zip codes, postal codes, race, gender, date of birth and place of birth. We consider this information private as this can easily be used to identify an individual.

In addition to this, geo-location information and data on the device e.g messages, emails, identifiers, and logs need to be protected while using modern apps.

With these key gaps in mind, we propose a privacy standard mobile application framework in an effort to improve privacy protection in current mobile applications. The following chapter elaborates on the contributions of this thesis as a means of tackling these gaps.

Chapter 3

Research Methodology

3.1 Introduction

This chapter explains the research method adopted in this thesis. We present the research methodology applied and the procedures we undertook to accomplish all the tasks in the research. We differentiate the types of research methods used which include qualitative research methods and quantitative research methods. Qualitative research involves performing research using empirical methods that are not in the form of numbers e.g. by using interviews and questionnaires. It assumes a dynamic and negotiated reality where data is collected by performing interviews and through participant observations. On the other hand, quantitative research assumes a fixed and measurable reality which is gained by collecting data through measuring things. In quantitative research, data is analysed using numerical comparisons and statistical inferences [82], [83]. This research involved using both a qualitative and a quantitative approach. In the following section we present details of the research design and its structure.

Research Methodology: A research methodology is the process of investigating research topics and is further referred to as a theoretical underpinning of the research. It is a way of defining the research activities involved, the procedures taken to accomplish the research, determining the elements the research is based on, defining the scientifically adopted models, designs and tools.[84]. Research methodologies are generally classified

as quantitative and qualitative research. However, they can be further classified into descriptive and analytical research methods. Descriptive research methods involve the use of surveys and analytical research uses facts to analyse a problem in order to make a critical evaluation of the material involved in the research. Other types of research such as exploratory research on the other hand explore the research area and do not attempt to attain a conclusive answer to the research questions involved [85], [83]. Thus, choosing a research method should be taken with much care as there are many factors to take into account especially in how the studies should be evaluated and presented to the general audience.

3.2 Research background and research questions.

This section presents an overview of the research background detailing the problem domain and the need for performing the research. It also elaborates on the importance of users knowing the privacy measures integrated in the apps e.g in the form of privacy policies and their rights to using applications that are not privacy invasive.

3.2.1 Mobile application selection and analysis

The use of mobile applications has steadily increased since the inception of the Android Google Play Store and the App store. At the time of writing this thesis, Android Google play store was reported to have 2.6 million apps in March 2018 and App store 2.0 million apps. The overwhelming number of applications commonly known as apps, on the market comes with a lot of challenges faced both on the developer side and the consumer side. As we have reported in chapter 1, there have been reports of security and privacy breaches which have constantly increased as the number of mobile applications increase. A study made by the privacy body based in Washington DC, the *Future of Privacy Forum* in 2012 found that 48% of free mobile applications had in-app access to a privacy policy while 32% of paid applications did have one [86]. With these findings in mind, it is eminent that at least half of the users' of mobile applications do not know that privacy policies do exist. For this reason most of the users do not know why privacy policies are required

and what they are used for. In previous years, Google play store has however, enacted privacy policy requirements for developers which demand that all apps should contain a privacy policy that is easily accessible in the application by users. This was after the California Online Privacy Protection Act (CalOPPA) was enforced in 2012. CalOPPA applies to any person who develops mobile applications that collects Personal Identifiable Information (PII). Developers should include a privacy policy whenever they share or use PII and clearly state how they are going to use the data collected.

In addition to these laws, the new GDPR requires that users are well informed of the data collected about them and how the data is processed. Privacy policies among others must include:

- The identity of the data collector (address, name of the company) has to be given.
- The type of data the mobile application collects.
- Why and how the data is to be processed.
- The identity of the third parties the data is shared with.
- Users should be able to request for the data collected about them.
- Users should be able to delete their data.

In an attempt to reach a more nuanced understanding of how current mobile applications are designed and how privacy policies are integrated in mobile applications we examined if the applications have in-app privacy policies and if the privacy policies are on the company websites. We further examined which permissions the application uses The major key questions we investigate in the course of the methodology are as follows:

- What are the current trends of privacy preserving procedures integrated in mobile applications in current App stores?
- Are the privacy policies included in both the companies websites and the mobile applications?

- Are the in-app privacy policies readable and easy to understand for users?
- Are users explicitly told of which data is being collected and for what purpose?
- Are the mobile application permissions in the app stores easily found and easy to understand by users?
- Are mobile application permissions necessary for the functionality they are requested for or not?

To answer the above questions we focused our investigation on the Google Play Store. We considered Google Play Store because its apps can easily be accessed and are more in number compared to those in App store. Google Play Store is a digital store where users can search and install mobile applications. The applications are installed on Android devices. Android is an Open source software framework which provides APIs based on the Android Software Development Kit. Android is community based, which means that developers are allowed to develop applications in the Java programming language and publish them in the store. To investigate the current state-of-the-art in mobile applications, we examine 50 most popular mobile applications. We focused on two major domains, the eHealth domain and the Telematics domain. The data was collected between December 2018 and June 2019. We used Google play store to determine the type of permissions the apps use.

3.3 Research methods used in this study

3.3.1 Qualitative research methods - documentary analysis

The first qualitative research method used was that of a documentary analysis. Documentary analysis involves the study of written material in form of documents, reports, peer reviewed publications like journals, research papers, newspapers and media online. It focuses on what is reported in these documents. The analysis of the documents can be done electronically using content analysis software, or text data mining tools, or manually focusing on the words and phrases in the documents. The type of document analysis used

in this research was done by undertaking a literature review which has been presented in chapter 1. We investigated media reports both online and in newspapers to study the problem domain. We searched in online data bases e.g IEEE, ACM, Springer, Elsevier and used key words such as *privacy in mobile apps*, *privacy protection in mobile apps*, *data breaches in android apps* to get an insight of the problem domain.

3.3.2 Exploratory research and analysis - of mobile apps

The next research method used was that of an exploratory study in eHealth and Telematics apps. This study was investigative and did not involve any testing and evaluation rather was conducted to determine the extent at which mobile apps do integrate privacy preserving methods in the application. We looked at how the apps were developed, which features they had to offer, checked the company websites if they did have privacy policies. We further investigated the type of data that was collected both in the app and what was stated on the company sites. In the following we present how the exploratory research of the eHealth mobile applications was done.

3.3.2.1 Privacy requirements analysis in mobile apps

We subdivided eHealth apps into six categories; Fitness apps, Cardiology apps, Diabetes apps, Weightloss apps, Depression apps and Physician apps. We searched for the apps in Google Play store according to the above named categories. Our aim was to determine the following type of data collected:

- Which permissions does the app have access to?
- Are the permissions requested required for app functionality?
- Does the app have a privacy policy in Google Play Store?
- Does the app have an in-app privacy policy?
- Does the app have a privacy policy on the companies website?
- Does the privacy policy contain the details of the data that it collects?

- Does the privacy policy contain the contact of the developer?
- Does the app inform users of the medical data it collects?
- Are users informed of the data shared to third-party service providers?

The chosen apps were subdivided amongst the above mentioned categories. The apps are a combination of health improving apps and health information apps. They are used to enhance modern health care management for patients and health care providers. We further examined, patient monitoring apps which gather patient information and electronic health care records. The apps were of interest because they collect a lot of health related personal information ranging from patients genetic conditions required for weightloss, to hereditary conditions like diabetics, and mental health conditions like depression. Such personal data is very sensitive and requires a high level of privacy during app development and in operation. The investigated apps are open source and do not require stringent medical observation from medical personnel, so no special medical guidelines were required. We focus on eHealth fitness and Telematics mobile applications because of their constant evolving nature. These two app domains collect a lot of sensitive data without going through stringent controls that are required to protect user data. For these reasons we found it necessary to investigate these domains in an effort to improve user privacy in current mobile applications.

3.3.2.2 Data extraction in mobile applications:

The initial step of the data extraction was to download the apps from Google Play Store on a Motorola XT1092 phone running Android version 8.0. We analyzed the information given in the store about the app. This was done in two phases; (1)First, we examined the permissions required while downloading the app and (2) the second phase was to examine the privacy policies. We transcribed in detail which parameters were indicated in the permissions and the privacy policies. All information that was requested by the companies during sign-up and that was required in the privacy policies was recorded. Key parameters especially PII details like; name, gender, email addresses, phone numbers to financial data

such as credit card numbers, employer were all recorded for privacy policies on company websites and within the apps. Table 4.15 shows the common permissions which were recorded in the various apps we examined. Generally apps have different features as they provide various services to customers, we recorded all the optional information that apps required, for example Lose it! which is a weightlos app requires that is customers give up a saliva sample to determine their DNA for weight loss purposes.

3.3.2.3 Permissions required in mobile applications

We manually recorded the permissions the apps required in the Google play store. The permissions required after logging in the app were further examined and compared to those in the store. We examined all permissions of the application as presented in the Google Play Store and within the app. We transcribed all permissions as seen in Table 4.16 for the Nike training app and privacy policies as shown in Table 4.17 for the Depression CBT app and Table 4.18 for MySugr app, to determine the type of permissions and privacy policies the application requires. The results were recorded for each app and observed that some apps requested for permissions which were not used for application functionality. We transcribed all the details regarding the permissions and attempted to answer the following key questions;

- What are permissions used for?
- What type of permissions are used for the apps?
- What permissions are required for the app to function?
- What permissions are not required for the app to function well?

3.3.2.4 In-app privacy policies in mobile applications

The privacy policy given in the store was examined and screen shots were taken in an effort to determine which data the app collects, is mentioned in the privacy policies. All information about the collected data in the privacy policy was transcribed. After downloading the apps, accounts were created for each app and the information requested

in the app was recorded. This information included if the app requested for personal information like the name, email, address or a login using social networks like Facebook, or the app required payment information like credit card details. Furthermore, we recorded if the app had a privacy policy during log on or not. This data enabled us to compare both privacy policies in the app and that in the Google Play Store to determine any inconsistencies of personal data collected by the apps. Next we examined the privacy policies in the company websites to determine if they differed from those in the app and the store.

3.3.3 Quantitative approach - Using Privacy scores

The next approach used was a quantitative approach where we applied privacy scores to determine privacy integration by finding out which apps collected more private data in comparison to others. The aim of this approach was to investigate if the data collected in the apps was in sync with the functionality that the apps offer and if it was privacy invasive. The scores were obtained by identifying key words of what type of data the apps collect. These were transcribed to generate scores which included, a data collection score, a functionality score and and a privacy policy score. The data collection score was computed by identifying which type of data is collected, the functionality score computes the type of functionality and features that the apps offer while the privacy policy score computes the amount of privacy policy measures implemented in the app.

3.3.4 Qualitative analysis - Conceptual framework design

A conceptual framework is required to synthesise the literature and provide the answers to the knowledge gap in the literature. We identified that there are limitations in the design of privacy aware applications as companies did not fully integrate privacy in current mobile applications. In an attempt to resolve the privacy limitations, we took a qualitative approach and developed a privacy aware trade-off analysis framework (TRANK) aimed for system designers, developers and privacy engineers to design mobile applications that integrate privacy at the initial stages of development. The framework entails a privacy

goals component, a privacy goals trade-off analysis component, a privacy requirements component, a privacy and functionality trade-off analysis component and a privacy requirements management component. These components are used to integrate a privacy development life-cycle during application development.

We further used a prototyping methodology to create prototypes which were used to evaluate TRANK and as a proof of concept for the framework. Two frameworks were designed; a Telematics app prototype and an eHealth fitness app prototype.

Finally the prototypes were evaluated using interviews and questionnaires. In order to assess how users perceived the applications we carried out two interviews based on questionnaires. The first interview was a privacy based interview which aimed at asking participants how they perceived the privacy features integrated in the apps. We recruited participants who were active app users and asked their opinions on how the privacy integrated in the apps should be enhanced and what their opinion on privacy implementation in current apps was.

The second interview was a usability interview. This aimed at asking participants if the apps were convenient to use in terms of usability and what they would like to improve.

3.4 Summary

This section has presented the methodology used in this thesis. We present two major research methodologies namely the qualitative and quantitative research methodologies that we use in the course of this research. We specify how the research processes in each methodology were carried out beginning with a qualitative approach using a documentary analysis which involved using a literature review to define the problem scope. In addition, an exploratory research method was undertaken to investigate the challenges of privacy preservation in mobile applications. A quantitative research methodology was employed to analyse the collected data in the investigated mobile apps using privacy scores. Further qualitative methods employed include; (1) the design of a conceptual privacy aware framework which serves as a tool for privacy engineers and designers to develop privacy aware applications (2) a prototyping methodology used to design the prototypes which

serve as a proof of concept of the research undertaken in this thesis and lastly surveys were carried out to evaluate the prototypes. The surveys included a privacy related user survey and a usability survey.

Chapter 4

Contributions

4.1 Empirical study 1 - Exploratory Empirical study

This thesis presents three empirical studies which investigate the gaps related to designing privacy aware mobile applications. The first contribution of this thesis is the exploration of modern telematics mobile apps which involves an investigation in the type of data the apps collect and the privacy methodologies integrated in the telematics applications. We further carried out a study in modern eHealth apps to similarly determine what type of data is collected and how the apps integrate privacy preserving methods. In an effort to protect user privacy especially in the eHealth domain, where sensitive medical data is gathered and processed to provide health related services e.g Diabetics management, mental health management and weight loss management among other services, it is important that medical data is processed in a proper manner. In this section we present the approach used in performing this study and our key findings of the study.

Privacy in evolving mobile applications

There has been a surge in the deployment of telematics mobile applications in recent years. One of the promising revenue streams in the telematic ecosystem is the use of diagnostic data to provide insurance-relevant information. The use of telematic insurance requires the collection of vast amounts of vehicular data. Some of these data are clearly indicated on the websites of insurance companies. However, most insurance companies do

not inform users about the data being collected and the purpose of doing so. We report on the results of an exploratory study on the top 5 telematic insurance or black-box insurance companies in the UK. We analyse the data collected by telematic insurance companies, compare the privacy policies given on companies' websites, model privacy requirements and offer insights into privacy-requirements engineering in telematics insurance applications.

4.2 Privacy in Telematics mobile applications exploratory study

There has been an increase in telematics mobile applications deployed in modern vehicles in recent years. Telematics mobile applications are also referred to as V2X Telematics applications. V2X applications are applications that are used to communicate between Vehicle-to-Vehicle (V2V) or Vehicle-to-Infrastructure (V2I) systems. Examples of V2X applications include emergency call systems, vehicle theft-recovery systems, fleet-management systems and remote vehicle diagnostics. Among the emerging V2X technologies is the use of telematic data to advise insurance premiums to users. This is referred to as telematic insurance or commonly known as black-box insurance. Insurance companies are offering black-box insurance policies in addition to traditional premium policies. Black-box policies are aimed at 'young drivers' between 17 and 25 years of age. Users are required to install a 'black box' which is a telematics device that constantly records and monitors vehicular data on how, when and where the user drives.

With the help of this telematic information, insurance companies are able to accumulate a vast amount of personal data. However, the privacy impact of collecting telematic data and the role of privacy requirements engineering in designing telematic insurance software have remained unstudied to date.

A number of major questions remain unanswered and act as the basis of this study:

- Are users aware of the privacy concerns related to the adoption of systems provided?
- Are users explicitly informed about the data being collected so that they are able to opt-in or opt-out of applications?

- Do the privacy policies on companies' websites include information about the data being collected, how they are collected and what they are used for?

We performed an exploratory study on telematic insurance providers in the UK in order to find possible answers to these questions and to determine how privacy-requirements engineering in telematic insurance systems can be modelled and implemented. This section presents the initial results of our findings and highlights the gaps and challenges encountered when engineering Telematics mobile applications privacy requirements. The novel contributions of this study are threefold:

1. We present insights into the information collected in telematic (black-box) insurance.
2. We highlight the privacy risks involved in using telematic insurance and compare the data collected by the system to those included in privacy policies. Our analysis highlights some potential mismatches between privacy policies and the data collected (reflecting similar gaps identified by previous studies of privacy in online social networks [87]).
3. We apply an existing privacy requirements modelling framework, LINDDUN [33], to derive privacy requirements for telematics insurance and discuss the potential challenges in realising these requirements in practice.

Privacy implementation in Telematics applications

Privacy plays a key role in the design and implementation of vehicular communication networks (VCNs). To this effect, several V2X research projects have introduced privacy requirements. One of the key VCN projects that focuses on privacy in V2X networks is the European Union Project: Preparing Secure Vehicle-to-X- Communication Systems (PRESERVE). The PRESERVE project aims to protect users' privacy by introducing an integrated V2X Security Architecture (VSA) that incorporates pseudonyms as a mechanism to improve privacy [88]. Similarly, the PRivacy Enabled Capability In Co-Operative Systems and Safety Applications (PRECIOSA) project aims to integrate privacy-preserving mechanisms in VCNs and to formulate regulations and privacy guidelines for

protecting data and location privacy in vehicular cooperative systems [89]. These works serve as guidelines for modelling telematic privacy requirements in our study.

In the following we present prior research in two major categories, that based on anonymisation and data obfuscation and that which focuses on location privacy.

4.2.1 Anonymising and data obfuscation

Data anonymisation and obfuscation in VCNs is mainly done by using pseudonyms. Pseudonyms are a set of short-term identifiers that ensure the privacy of users by hiding the real identities of senders. The adoption of pseudonyms has been extensively studied. For example, Chaum et al. introduce the use of digital pseudonyms as a means of providing privacy by hiding the identity of a user transferring information from one source to its destination [90]. To enable the unlinkability of vehicles, the use of changing pseudonyms at a given time and period has been proposed and pseudonym schemes have been extensively compared and analysed to determine the best strategies [91]. However, the use of pseudonyms has been challenged by Wiedersheim et al., as there remains the possibility of linking vehicles to their users, depending on location information and the level of granularity [92].

4.2.2 Location privacy

Location information remains a major concern in the implementation of telematics mobile applications and has been extensively studied, in detail, to demonstrate privacy breaches that occur using vehicle-profiling. Ma et al. highlight privacy concerns in the continuous accumulation of location-driving data which may be used to determine a user's neighbourhood over a long period of time [93]. Hoh et al. show how location privacy in traffic-monitoring systems that use GPS equipment can reveal the position and location to the service provider's traffic-monitoring servers. In [94], the authors propose the use of multiple pseudonyms as a means of protecting location privacy and maintain that some actors may not be willing to change pseudonyms.

Research in both of these categories plays a key role in enhancing privacy protection

in telematics mobile applications systems and emphasises the importance of protecting personal data. While the above works emphasise privacy protection in the V2X domain in general, this thesis focuses in particular on protecting privacy in mobile applications, especially the concern for explicitly raising privacy awareness and the protection of personal information collected by telematic insurance providers.

4.2.3 Data acquisition

We analysed the top 5 telematic (black-box) insurance companies in the UK in order to determine the types of personal, telematic data or driving data aggregated and how they are used. The information collected was found on company home pages at both main URLs and privacy-policy URLs. Data acquisition was performed by taking snapshots of the data collected by the black-boxes as stated on company sites. Collected data were taken at face value from the websites. The process of data collection was performed between November 2016 and April 2017. Initially, we recorded the information presented on companies' sites about the data collected by their telematic devices (black boxes). We then recorded the information presented at companies' privacy-policy URLs. The final step was to contrast the information presented about data collection at company URLs with that presented at privacy policy URLs to identify potential mismatches. The following steps outline our data-acquisition process:

- **Step 1**: (a) Search for the top 5 telematic insurance-provider sites. (b) Check if each site has a page explaining how black-box insurance works. (c) Elicit results by marking sections advising the data collected by telematic devices. (d) Check if sites contain information collected in case of an accident, take screenshots and compare sites. (e) Note down the results in a comparison table.

- **Step 2**: (a) Check if telematic insurance provider sites have a privacy-policy site. (b) Repeat substeps (b), (c) and (d) in Step 1 above. (c) Check if the privacy-policy site has information about third-party service providers. (d) Check if the site has information about where data are stored and processed by third-party service providers (EU, Non EU). (e) Check if sites include information about what happens

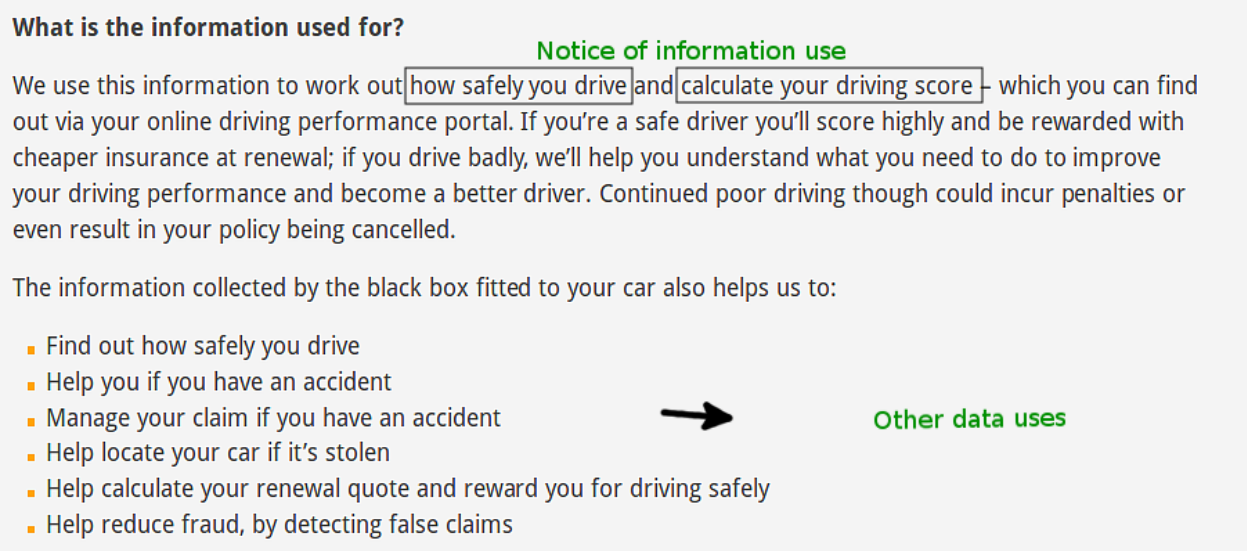


Figure 4.1: Example of information use in black-box insurance

to data after users have opted out of policies. (f) Note down the results in a comparison table.

4.2.4 Search Results

This section presents the results of our study. Figure 4.1 provides an example of information collection in black-box insurance given on one of the company sites [95]. Next, we outline the findings from our study in detail. Our results are based on three major observations:

1. Data collected by telematic devices on a daily basis.
2. Data collected and further processed by third-party service providers.
3. Data collected even when a user has opted out of telematic insurance services.

A. ***Data collected daily by the telematic device.***

Most traditional insurance providers offer discounts on premiums based on no claims. In contrast, with telematic insurance, telematic data are used to determine a driver's premium. Generally, there are two major types of data collected by telematic insurers: personal data, i.e. information filled in on quotation forms, e.g. name, gender, address and age, which is used to file for claims. The other type

Data Collected by Telematic devices given on websites

	Think	Marmalade	SmartMiles	Insurethebox	DrivePlus
Data Collected					
Time driven	X		X	X	
Duration of driving	X		X		
Speed	X	X	X	X	
Acceleration	X	X	X	X	
Brake	X	X	X	X	
Breaks taken	X			X	
Cornering	X	X			
Number of miles	X			X	
Number of journeys	X				
Roads you drive on	X		X		
Motorway miles				X	
Vehicle location	X		X	X	
Accidents					
Time and place	X				
Force of impact	X				
Direction of travel	X				
Speed before impact	X				
Speed after impact	X				
Privacy policies					
Data in privacy policy				X	partial
Use in privacy policy				X	partial
How Data is used	X		X	X	partial
How Data is shared			X	X	partial

Figure 4.2: Data collection provided on the companies' web sites.

of information collected is telematic data (data collected via black boxes). The black box contains four major components that help in acquiring telematic data: (1) a GPS satellite receiver which aids in locating the vehicle, wherever it is; (2) an accelerometer which detects whether sharp cornering or a collision has occurred; (3) a SIM card which sends collected data to back-end databases; and (4) customised insurance software which analyses and sends collected information to insurance back-end servers.

Figure 4.2 gives a brief report of the details we found about the data collected by telematic insurance providers, which are used to calculate monthly premiums. A letter X in the table denotes the type of data reported as being collected.

There are three key types of driving information collected by telematic devices: acceleration, speed and cornering. The location of a vehicle is tracked via a GPS module which is used to find the vehicle in case of theft. A motion sensor is installed to detect if a vehicle has had an impact in case of an accident or heavy braking.

The speed at which a vehicle is driving is tracked to determine whether its user is driving within legal speed limits. The duration for which a driver has been driving is recorded to determine if the driver has been driving for long hours. Telematic insurance provider further record if one takes breaks while driving long distances or when one is driving on a motorway (highway). The number of miles (and motorway miles) driven, the roads that are driven on, if one is driving at night or during the day are all recorded and some of this information is provided to an online portal for users to get an estimation of the monthly insurance premium used for billing and payment purposes beforehand.

For each of the insurance providers in our study, we collected telematic data directly from the information provided on their webpages [95] [96] [97] [98] [99]. Most of this information is not given directly on a website but has to be searched for at external URLs indicated, for example as *"How does black-box insurance work?", "What is telematics?", "What is black-box insurance?"*. We observed that some of the telematic insurance providers explicitly explained what telematic insurance is and gave the type of telematic data that the device collects and how they are used [95] [98]. However, for some sites, a comparison of the data given on the webpage differed from that given in the privacy policy. In detail, we found that some companies only gave a general list of data collected, with statements such as *"Our black box only measures four aspects of your driving - the ABC's"*, meaning that they collect data about acceleration, brakes, cornering and speed with no further explanation of what exactly the data are used for [96].

B. ***Data collected and processed by third-party service providers.***

Telematic data are initially collected via a vehicle's OBD-II port. This aggregated information is then transmitted to the insurance company over a wireless phone network. To enable proper processing of collected telematic data, insurers employ third-party service providers to deliver services, e.g. calculating premiums or detecting fraud. Third-party service providers play a big role in the development and implementation of telematic insurance, as collected telematic data and personal in-

formation are constantly shared between insurers and among other service providers. Examples of such service providers are: *credit-reference agencies which aid in checking users' identity and credit status, fraud-prevention agencies, third-party libraries, driving-licence authorities, research agencies that analyse insurance products, online databases, regulators and law-enforcement agencies.* Although telematic data and personal information are obtained or transferred to third-party service providers, most telematic insurers do not include this in their privacy policies. The few sites that included the type of information they share with other service providers were not precise, using phrases such as: *"we will only share your driving information with our trusted business partners".* The sharing of personal information is not included in privacy policies, instead rather general information is provided, e.g. *"we will release personal information to other insurers, or agents providing services to us on your behalf"* [95]. In addition to this, some telematic insurance providers mention that *"the data we collect from you may be stored outside the European Economic Area (EEA)"* [97] [98]. This not only means that personal data are transferred to other service providers outside Europe but also indicates that the insurance companies are not ready to mention the countries to which the information is transferred. Finally, some sites mentioned that sensitive information, e.g. health, race, religion and previous criminal convictions, is used when determining user policies. This clearly indicates that users have to be explicitly informed about such information before they opt for using telematic insurance applications.

C. ***Data collected after users have opted-out of telematic insurance services.***

According to our findings, most telematic insurance companies did not give information about how telematic data are used when a user opts out of a policy. Opting in and out of a policy means that the user is free to choose and buy (opt in to) the policy or cancel (opt out of) the policy. Some telematic insurance companies state that *"we will stop recording your information if you sell your car or if your policy is cancelled"* [98]. However, we found it contradictory when the same insurance companies claim that they will use information to *"carry out research and analysis*

about our products and services" even when a policy is cancelled. Therefore, it is not clear what telematic data are used for after users sell their cars or cancel their policies.

4.2.5 Deriving Telematics mobile applications privacy requirements

From the search results obtained in section 4.3.4, we observed that privacy implementations by current telematic insurance providers are inconsistent and unsatisfactory. One of the major research gaps in their design is the lack of adequate privacy requirements in engineering models specific to telematic systems which have to be employed in the initial stages for stringent privacy-aware application development. Although the V2X community has raised concerns that question privacy non-compliance in the majority of telematic insurance systems on the market today, little has been done to model privacy requirements aimed at telematic insurance. It has not yet been established how telematic insurance should be adopted while offering functional system requirements without compromising privacy at the same time. We, therefore, apply one of the general approaches for modelling privacy requirements, LINDDUN [33], to derive privacy requirements for telematic insurance and analyse the challenges in realising these.

The first step in the LINDDUN methodology is to create a data flow diagram (DFD). This is done by generating a graphical representation of the system to be modelled in terms of how data are transferred between its components. A DFD contains 4 major elements: (1) external elements that use the system (e.g. users, external applications and services); (2) processes that are run on the system which include run programs, procedures and system functionality processes; (3) data flows which represent data transfer from one component in the system to another; (4) databases or data stores which store collected vehicular transactional or logical data.

Creation of a Data flow diagram

The abstract data-flow diagram in Figure 5.7 describes how data are exchanged in telematic insurance applications, e.g. when information is requested by a driver to show the current status of their driving behaviour or insurance premium online. A telematic

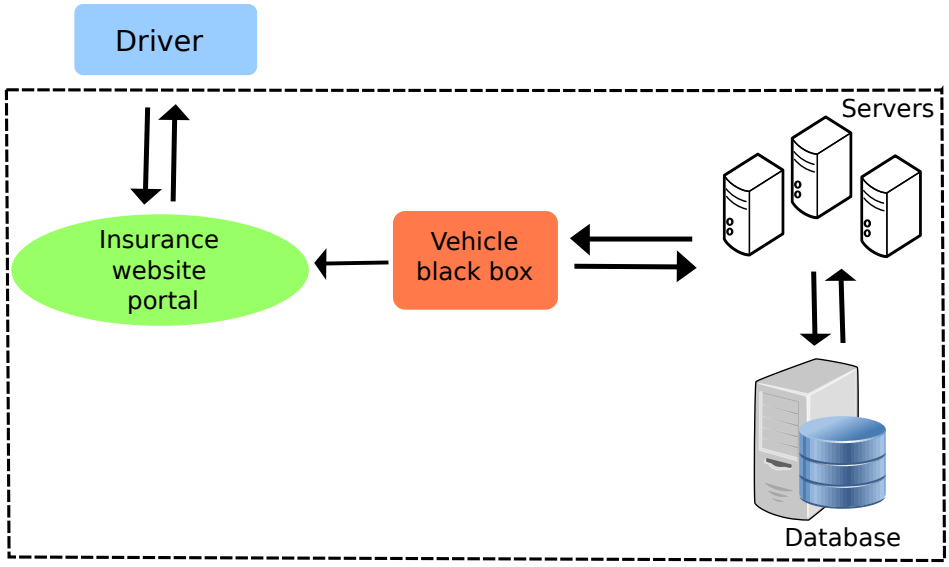


Figure 4.3: Telematics insurance data flow diagram

application is composed of four main processing components: (1) an insurance web portal, which drivers use to enquire about their premium status; (2) a vehicle black box, used to collect telematic data; (3) back-end servers, used to process collected data and (4) databases, used to store telematic data. These components have to be protected against malicious intruders that could compromise user privacy in the system.

Mapping of DFD to potential privacy threats

The next step is to map the DFD to potential threats that may arise when using a telematic insurance system. The mapping involves identifying threats that may affect each component of the system. This involves distinguishing the type of threats that are applicable in data transfers from users to the web portal, telematic devices to servers, and servers to back-end databases. From our analysis, each DFD component is susceptible to privacy leakages according to the following LINDDUN privacy categories:

1. Linkability (L).
2. Identifiability (I).
3. Non-repudiation (N).
4. Detectability (D).

Table 4.1: Mapping of Data flow diagram to potential threats

Mapping of DFD to privacy threats								
	Privacy threat target	L	I	N	D	D	U	N
Telematic Data Store	Telematic insurance DB	•	•	•	•	•		•
Telematic Data Flow	Driver - portal flow	•	•	•	•	•		•
	Portal - driver flow	•	•	•	•	•		•
	Portal - vehicle flow	•	•	•	•	•		•
	Vehicle - portal flow	•	•	•	•	•		•
	Vehicle - server flow	•	•	•	•	•		•
	Server - vehicle flow	•	•	•	•	•		•
	Server - database flow	•	•	•	•	•		•
	Database - server flow	•	•	•	•	•		•
Telematic processes	Telematic insurance portal	•	•	•	•	•		•
	Manage vehicle black box	•	•	•	•	•		•
	Manage servers	•	•	•	•	•		•
Entity	Driver	•	•				•	

Table 4.2: Telematic threats modeling

Description T1	Direct vehicle identification which may be linked to users.
Attack process	Adversaries who gain access to the system may identify users and their vehicles.
Result	Access to the system to identify which vehicles belong to a particular driver, exposing their identity.
Description T2	Vehicle and driver location identification.
Attack process	Vehicles are located using GPS receivers and geo-location data.
Result	Location identification of drivers and vehicles.
Description T3	Information-sharing of private data to third-party providers.
Attack process	Telematic insurance providers sharing or selling data to third-party service providers.
Result	Information disclosure and misuse by third-party service providers.

5. Information disclosure (D).

6. Content unawareness (U).

7. Consent or policy non-compliance (N).

Table 4.1 shows the mapping of a data-flow diagram to potential privacy threats. A black circle symbol • denotes possible breaches of a privacy category.

Generating telematic insurance privacy threats

After mapping the DFD to potential threats, we derived potential privacy threats in relation to DFD elements. Table 4.2 illustrates threats that may arise when using telematic insurance systems. It describes privacy breaches that may be used to compromise the system. The list includes a description of the attack, the type of attacker, the attack process that may be used and finally the result of the aforementioned attack. Threats are

Table 4.3: Deriving privacy requirements from threats

From threats to privacy requirements	
Telematic insurance threats	Privacy requirements
Linkability of users to telematic insurance DB	Unlinkability of users to database
Linkability of users to data flows	Unlinkability of users to data flows
Linkability of users to telematic processes	Unlinkability of users to system processes
Linkability of external entities, e.g a driver,	Unlinkability of external entities connecting to systems
Identifiability of users to telematic insurance DB	Pseudonymity or anonymity of users in DB
Identifiability of users to data flows	Pseudonymity or anonymity of users in data streams
Identifiability of users to telematic processes	Pseudonymity or anonymity of users in systems processes
Non-repudiation of users sending data streams	Acknowlegement of users using system processes
Non-repudiation of users using system processes	Acknowlegement of users using system processes
Detectability of users connecting to DB	Unobservability of users connecting to DB
Detectability of users sending data streams	Unobservability of users sending data streams
Detectability of users using system processes	Unobservability of users using systems processes
Information disclosure of users connecting to DB	Confidentiality of users connecting to DB
Information disclosure of users sending data streams	Confidentiality of users sending data streams
Information disclosure of users using system processes	Confidentiality of users using system processes
Unawareness of users about content given in systems	User awareness of content given to systems
Non-compliance of systems with policies and regulations	Compliance with policies and regulations in systems

denoted by the letter "T" and are numbered in series T^1 - T^n depending on the system to be modelled and the number of threats derived. Please note that the list is not exhaustive, rather, only a representative fragment is presented in this thesis.

Generating privacy requirements from threats

After distinguishing potential threats, privacy requirements are then derived on the basis of threats as illustrated in Table 4.3. Here, we followed the LINDDUN threat categories. Please note that these requirements are not exhaustive, rather they should be seen as examples of privacy requirements for telematic insurance applications based on the methodology described in [33] and are subject to revision in case the system changes or a different type of system altogether has to be modelled.

The major privacy-requirement engineering implication of this study is that there are major potential privacy threats, and hence requirements. Therefore, we need systematic ways of capturing these requirements and designing them into privacy-compliant telematic insurance systems. We discuss some insights into this next in our analysis.

4.2.6 Privacy challenges in Telematics mobile applications

From our analysis we found LINDDUN is a well-structured framework to model privacy requirements. Table 4.4 gives a general overview of the advantages and drawbacks of

Table 4.4: Advantages and drawbacks of LINDDUN privacy-requirements modelling

Advantages	
	- Provides a detailed structure to model privacy requirements. - Entails a privacy-by-design approach. - Models risks by deriving privacy-enhancing technologies from elicited privacy threats.
Disadvantages	
	- LINDDUN privacy-requirements modelling may be at too high a level to model large systems, and depends highly on the depth of the DFD. - The final product might not meet user requirements due to ambiguities in privacy terminology.

modelling telematic insurance using the LINDDUN framework.

A key challenge in realising privacy requirements derived using LINDDUN is that not all information in a telematics insurance application can be entirely unlinkable. Although total unlinkability and undetactability may be desired, applications have to depend on certain information, e.g. personal information, location information or driving behaviour, for proper functionality. Therefore, complete anonymisation of personal credentials and driving data is very difficult, if not impossible. Thus, there should be a means of enabling such services while at the same time taking into consideration users' privacy demands and preferences.

Our results reveal that users are not explicitly informed about the information collected by telematic devices (black boxes) in either privacy policies or on companies' websites. Telematic insurance providers must employ better means of informing users about the information they collect, how they collect it and what it is used for.

In order to fulfil this goal, employing privacy engineering, especially in the initial stages of development, can play a big role in the development of privacy-aware telematic applications. This involves designing stringent privacy policies that comply with privacy regulations used to control the access to and distribution of personal data in the proposed system, in combination with the system in operational use.

The design of such systems is complex. Further, the complexity of managing privacy requirements in telematics is due to the fact that telematic applications are composed of dynamic sensor applications with constantly changing data. This means that designing privacy requirements involves dynamic data, e.g. acceleration, braking or speed data generated by electronic control units (ECU). This immense volume of data cannot be

stored in vehicles, it has to be sent to third-party services for further processing, making it very challenging to design efficient and expandable privacy policies. Another key challenge faced is the use of pseudonyms in telematic applications. Despite substantial previous research in this area [91], there is still a need for improved pseudonym schemes for telematic insurance applications to ensure that both system functionality and privacy requirements are met.

Furthermore, most telematic insurance service providers in the UK market today offer privacy-invasive applications which gather all the telematic data used for the calculation of insurance premiums. Insurance providers should therefore invest in designing system models that do not compromise privacy by, for example, only partially collecting telematic data, such as odometer and mileage readings, used to calculate premiums. In this way, data collection is reduced while simultaneously maintaining system functionality.

Consequently, the solution to these challenges is we believe as follows: First, telematic insurance service providers must ensure that the information on company sites indicates the types of data collected by the telematic devices. In addition, the data collected by insurance providers should be anonymised but at the same time allow proper functionality of the system. Second, third-party service providers and the applications they provide should be clearly indicated to users before they choose to opt in to insurance policies. Third, telematic insurance providers must ensure that policy statements given on company sites are complete and transparent to users.

Lastly, we note that there may needs to be systematic approaches for keeping policies and functionality in sync, as proposed by, for example,[100] and [101].

Above all, there is a need for revised privacy regulations and policy enforcement in telematic insurance development and deployment. Most importantly, users should be clearly informed about the data collected and how they are used. In this way, they will be able to opt in or opt out of insurance premiums due to unfulfilled privacy requirements or evident privacy concerns when they wish to do so.

Conclusions

We observe that there has been a tremendous increase in the deployment of mobile

applications in recent years. Telematics, or black-box, insurance mobile applications is one of the most promoted applications because drivers (especially young drivers aged between 17 and 25 years) would like to reduce the cost of vehicle insurance. Unlike existing insurance policies where drivers pay per year (PPY), telematic insurance is charged according to a pay-per-mile (PPM) or pay-as-you-drive (PAYD) model. Much as this development is promising for young drivers, though it does come with various privacy concerns. Telematics insurance is implemented by collecting driving data using a telematic device (black box) connected to the vehicle's on-board diagnostics II (OBD-II) port, which constantly tracks the movement and driving behaviour of its users. This study aims to enlighten users about related privacy concerns to ensure that user privacy is protected. Via an exploratory study of the top 5 telematic insurance providers in the UK, we demonstrate the privacy protection gaps encountered in the data-acquisition and privacy policies of mobile applications and in particular, telematic application services. We derive privacy requirements for telematic insurance applications using the LINDDUN methodology and highlight its challenges as regards realising total unlinkability and unobservability in this particular domain.

The following section presents an exploratory study of eHealth mobile applications to further investigate the privacy protection measures that have been integrated in eHealth mobile applications and how data collection is handled in current eHealth apps. We elaborate on the studies undertaken using eHealth android based mobile apps.

4.3 Privacy in eHealth mobile applications exploratory study

eHealth mobile applications also known as mHealth or telehealth apps have been on the increase in recent years. eHealth apps aid in managing patient health to improve and ease self management of health related information and the general quality of life for patients. Examples of eHealth apps include Physician apps, Mental health apps, Cardilogy apps, Diabetes apps, clinical trial apps e.t.c. Among the evolving eHealth technologies is the use of fitness data to improve patient health. It is believed that physical inactivity is one of the main causes of diseases such as cancer, diabetes, hypertension, obesity and vascular

diseases in adults. Subsequently, medical bodies like the NHS and health professionals have warned that physical inactivity can be deadly as smoking as was published in the Lancet [102]. Examples of eHealth fitness apps are Sworkit, Noom, Strava e.t.c. With the help of these apps, users are able to monitor their body activities and improve their fitness.

However, the privacy impact of collecting fitness data and the role of privacy requirements engineering in designing eHealth application software faces a lot of privacy challenges as was depicted in the Strava Data Heat Maps incident [14]. In this case, privacy concerns were raised as users who ignored the off-by-default privacy settings exposed off their location details. Through this, the fitness tracker exposed exercise routes of military personnel and their exact location. This led to major concerns in the way eHealth apps in general integrate privacy in mobile application development.

A number of major questions remain unanswered and act as the basis of this study:

- Are users aware of the privacy concerns related to the adoption of the eHealth apps used?
- Are users explicitly informed about the private eHealth data being collected so that they are able to opt-in or opt-out of applications?
- Do the privacy policies on companies' websites include information about the data being collected, how they are collected and what they are used for?
- What measures have been put in place to protect eHealth data especially medical data, genetical data and location based data so that it does not leak to the public?

We performed an exploratory study on eHealth fitness apps in order to find possible answers to these questions and to determine how privacy-requirements engineering can be modelled and implemented. This section presents the initial results of our findings and highlights the gaps and challenges encountered when dealing with sensitive medical data. The novel contributions of this study are threefold:

1. We present insights into the information collected in eHealth fitness applications.

2. We highlight the privacy risks involved in using eHealth fitness applications and compare the data collected by the apps to those included in privacy policies. Our analysis highlights some potential mismatches between privacy policies and the data collected (reflecting similar gaps identified by previous studies of privacy in online social networks [87]).

3. We apply an existing privacy requirements modelling framework, LINDDUN [33], to derive privacy requirements for eHealth mobile applications and discuss the potential challenges in realising these requirements in practice.

Privacy implementation in eHealth fitness applications

Privacy implementation in eHealth mobile applications is very crucial as they collect a lot of sensitive personal information like, medical records, location data, genealogical data e.t.c. Several researchers have reported on privacy [103], [104], [105] in eHealth apps however to the best of our knowledge no work has been done to integrate privacy-by-design using a privacy trade-off analysis at default. In the following we present some of the solutions employed in protecting user privacy in medical applications.

4.3.1 Anonymising and data obfuscation in eHealth applications

Anonymizing medical data has tremendously advanced in the areas of Electronic Health care records (EHR). Previous research has focused on anonymizing EHR records mainly to enable medical research based on patients data without exposing patients details. Using cryptographic techniques to anonymize data has been proposed in [106] based on hierarchical identity-based cryptography. Other mechanisms like l-diversity and k-anonymity [107] have been proposed to integrate privacy in electronic medical records. However, these come with limitations such as high computational overhead [108]. We have further seen a lot of eHealth related breaches reported as existent anonymization methods are not sufficient enough in protecting patient data. Therefore, this calls for more research in the way eHealth mobile applications have to be designed to meet future privacy demands.

4.3.2 Location privacy in eHealth applications

Location information in eHealth mobile applications is applied e.g in portable devices such personalised wearables used to monitor patients and other eHealth monitoring systems like fitness applications. Several studies have been taken in the eHealth domain to investigate the impact of location based applications. These are used in providing health remote services e.g diabetes management [109] which include web based services and smart phone based eHealth systems that connect the physician and the patient. Location based devices are further used to monitor patient data and offer personalized feedback to patients . Other studies proposed location based solutions to monitor mental health disease [110] which involved determining a patients location in order to investigate the relationship between location, depression and anxiety. In recent years, eHealth in fitness applications has been on the surge with more than 300,000 apps in app stores with numbers increasing. Fitness apps are the major players in using location information to monitor users. Much as location information has proved useful in providing information about users location that is crucial in the development and implementation of the above named applications, its use comes with a lot of privacy concerns as depicted in [111]. Therefore, there is a need for designing privacy preserving eHealth location based applications that ensure that privacy protection of medical data especially when using location information is ensured. In the following we present on the data acquisition performed and the results of our study.

4.3.3 Data acquisition in eHealth mobile applications

We analysed the top 5 top eHealth fitness apps to determine the types of personal, location and medical data that is aggregated and how its used. We gathered information from eHealth fitness app websites and recorded URLs for both websites and privacy policies. Data acquisition was performed by taking snapshots of the data collected by the apps and in the privacy policies. Collected data was taken at face value from the websites. The process of data collection was performed between May 2017 and July 2018. Initially, we recorded the information presented on companies' sites about the data collected by the

apps. We then recorded the information presented on the companies' privacy-policy URLs. The final step was to contrast the information presented about data collection at company URLs with that presented at privacy policy URLs to identify potential mismatches. The following steps outline our data-acquisition process:

- **Step 1**: (a) Search for the eHealth fitness apps. (b) Check if each site has a page explaining how the fitness app works. (c) Elicit results by marking sections advising the data collected by the fitness apps. (d) Check if sites contain information collected in case medical data is leaked, take screenshots and compare sites. (e) Note down the results in a comparison table.

- **Step 2**: (a) Check if fitness app provider sites have a privacy-policy site. (b) Repeat substeps (b), (c) and (d) in Step 1 above. (c) Check if the privacy-policy site has information about third-party service providers and if any information is shared to other companies. (d) Check if the site has information about where data are stored and processed by third-party service providers (EU, Non EU). (e) Check if sites include information about what happens to data after users have opted out of policies. (f) Check if the sites inform users of the medical data that is collected. (g) Check if users are informed before data is transferred and how they are informed. (h) Note down the results in a comparison table.

4.3.4 Search Results for eHealth applications

This section presents the results of our study. Figure 4.4 provides an example of information collection in a fitness app given on one of the company sites. Figure 4.5 gives a brief report of the details we found about the data collected by eHealth fitness applications in various user groups. We outline the findings from our study in detail. Our results are based on three major observations:

1. Data collected by eHealth fitness applications based on a daily basis.

2. Data collected and further processed by third-party service providers.

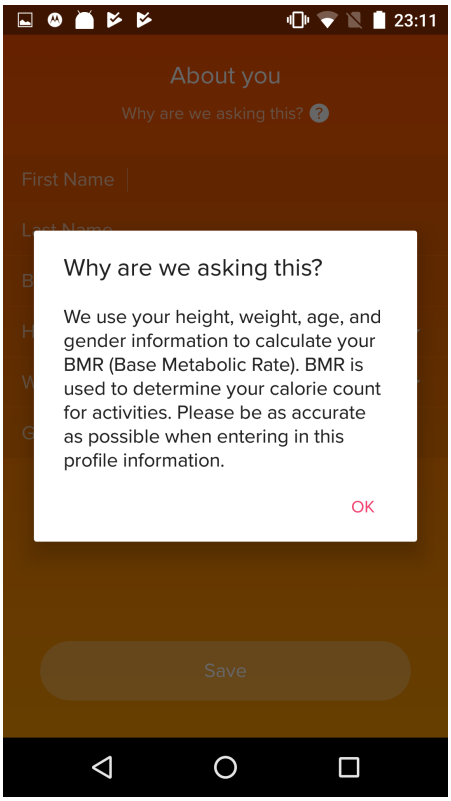


Figure 4.4: Data collected in fitness apps.

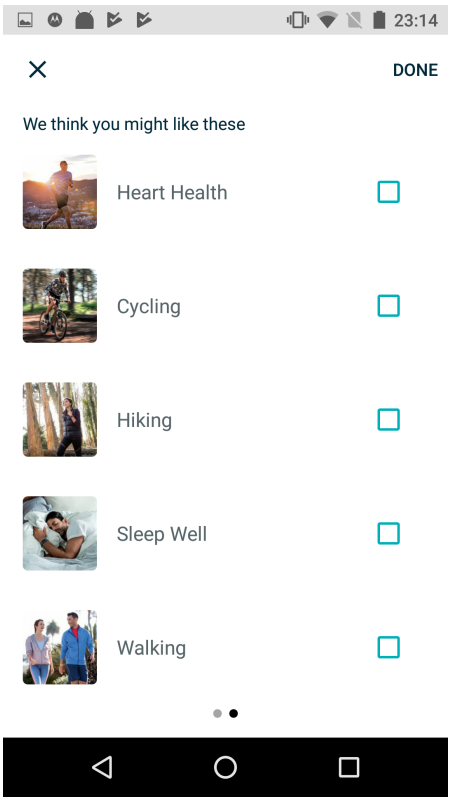


Figure 4.5: Data shared in fitness groups.

3. Data collected even when a user has opted out of the fitness app services.

4. How medical data is protected and stored after being transferred to companies.

A. ***Data collected daily by eHealth fitness apps.***

Data collection in fitness apps is increasing as fitness apps are on the rise in both Android and Iphone Operating systems due to the high level of demand. Current fitness apps are used to monitor users for their daily activities to improve physical inactivity. They monitor activities like, the heart rate, the route taken, the paces taken during physical activity. Fitness apps can be used to monitor diet by offering nutritional plans and dietary advise to users. We subdivided the type of data collected by fitness apps into 4 major categories. Personal data, Physical activity data, Nutritional data and Location data. Table 4.5 indicates some of the data collected and the activities performed. We note that fitness apps differ in the type of features they provide to their customers, therefore data collection varies according

Table 4.5: Data collected by fitness apps

Personal data	
	Name
	Age
	Gender
	Address
	Email
	Phone number
Physical data	
	Exercises done
	Daily steps
	Heart rate
Nutritional data	
	Meals taken
	Calories taken
Location data	
	GPS data

Table 4.6: Sensors used in fitness trackers

Sensors	**Measures**
Accelerometers	Acceleration
gyroscope	Orientation of an object in space
Barometric altimeters	Object elevation and altitude tracking
Compass	Cardinal direction
ECG sensors	Heart rate
Bioimpedance sensors	Pulse
Ambient light sensors	Light

to the features provided.

B. ***Data collected and processed by third-party service providers for eHealth fitness applications.***

Fitness tracking data is collected by individual app companies. This involves using sensors in the various tracking devices like smart watches and fitness trackers. The information is then processed by sensors. Table 4.6 shows some of the sensors used in fitness apps with the corresponding sensor data they collect. This data is then shared to third-party service providers to analyze the collected data and provide accurate services to the users. Several fitness apps have been reported to send user

data to social media platforms e.g Facebook [112]. This data is shared mainly for advertising and target advertising. This way companies can use it to advertise their products to the customers they presume will buy the products. Data shared ranges from daily fitness routines to more sensitive data like calendars and menstrual cycles. However, users are not informed of the companies which this data is shared to. The statements about data sharing are vague, and it is not clear which type of data is shared, if it is PII, sensor data, or motion data. The following excerpt 4.7 shows the information given about data sharing to third-parties found on Fitbits privacy policy.

We transfer information to our corporate affiliates, service providers, and other partners who process it for us, based on our instructions, and in compliance with this policy and any other appropriate confidentiality and security measures. These partners provide us with services globally, including for customer support, information technology, payments, sales, marketing, data analysis, research, and surveys.

Table 4.7: Fitbit privacy policy

This clearly shows that eHealth mobile app companies continue to share user data and the user is not aware of which impact it may have if this data is sold to insurance companies. Insurance companies may charge higher prices to customers that may have preexisting conditions or are physically inactive [113]. Therefore companies must stop sharing PII and sensitive medical data to third-party service providers for purposes of advertising to reduce privacy invasion.

4.3.5 Deriving eHealth mobile applications privacy requirements

The case study used in this study is that of a heart beat rate sensor as shown in Figure 4.6. We apply the LINDDUN methodology to generate privacy requirements, similar to the methodology of deriving telematics mobile applications privacy requirements described above.

Creation of a Data flow diagram

We initially created a Data Flow Diagram which shows the data flow between different

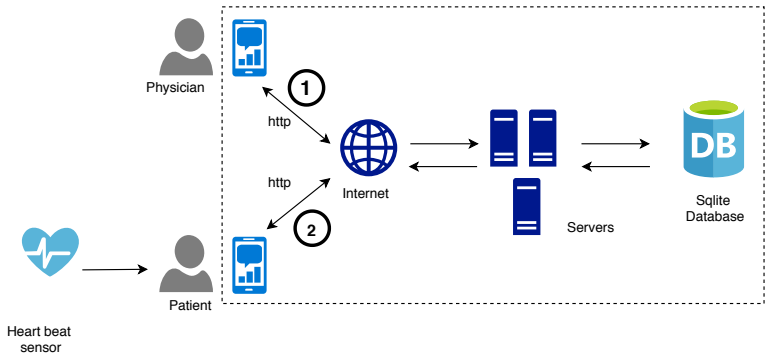


Figure 4.6: EHealth Heartrate monitor data flow diagram.

components of the applications as seen in Figure 5.6. The data flow diagram of heart rate monitor is composed of 5 major components (1) Heart beat sensors that measure a patients heartbeat (2) Patients (3) Physicians (4) Back end servers used to process the collected data (5) Databases used to store data. The components have to be protected such that eHealth data is not compromised.

Mapping of Heart beat monitor DFD to potential privacy threats

The data flow diagram is mapped to potential threats which may be used to attack the application. Each component is monitored to determine which threats may occur during the general data flow from each component to the other. According to LINDDUN, all components are prone to privacy threats and need to be analysed for potential privacy leakages based on the following privacy categories; Linkability (L), Identifiability (I), Non-repudiation(N), Detectability (D), Information disclosure (D), Content unwareness (U), and Consent or policy non-compliance (N). The mapping of the data flow diagram can be seen in Table 4.9.

Generate heart beat monitor privacy threats

Potential privacy threats to the eHealth heartbeat monitor were derived in relation to the generated data flow diagram. Table 4.8 illustrates some of the possible threats that may occur in an eHealth heart beat monitor application. This includes the attack, the attack process and the result of the attack. Corresponding threats are generated which further need to be modeled in order to find privacy solutions that best suit the threat.

Generate privacy requirements from threats

Table 4.8: eHealth heart beat monitor threats modeling

Description T1	Patient identification which may be linked to their illnesses.
Attack process	Attackers who may intercept the system and heart beat device may identify patients
Result	Compromised systems and access to patients sensitive medical data.
Description T2	Manipulation of the heart beat monitor and sensors.
Attack process	Devices are tampered with to produce incorrect values.
Result	Wrong measurements leading to false diagnosis.
Description T3	Information-sharing of private data to third-party providers.
Attack process	Ehealth service providers sharing or selling data to third-party service providers.
Result	Information disclosure and misuse by third-party service providers.

Table 4.9: Mapping of Heart beat Data flow diagram to potential threats

Mapping of Heart beat DFD to privacy threats								
	Privacy threat target	L	I	N	D	D	U	N
Heart beat Data Store	Heart beat DB	•	•	•	•	•		•
Heart beat Data Flow	Heart Monitor - patient flow	•	•	•	•	•		•
Heart beat processes	Patient - Server flow	•	•	•	•	•		•
	Server - Database flow	•	•	•	•	•		•
	Database -Physician flow	•	•	•	•	•		•
	Physician - Server flow	•	•	•	•	•		•
	Server - Patient flow	•	•	•	•	•		•
	Heart beat sensors	•	•	•	•	•		•
	Manage heart beat sensors	•	•	•	•	•		•
	Manage servers	•	•	•	•	•		•
Entity	Patient	•	•				•	
	Physician	•	•				•	

The next step after deriving privacy threats is to generate the anticipated privacy requirements. The privacy requirements are based on privacy categories; Linkability (L), Identifiability (I), Non-repudiation(N), Detectability (D), Information disclosure (D), Content unwareness (U), and Consent or policy non-compliance (N) as proposed by the LINDDUN methodology. Table 4.10 illustrates how privacy requirements are derived from threats based on the privacy categories. In the following we elaborate on some of the privacy challenges faced in eHealth mobile applications.

4.3.6 Privacy challenges in eHealth mobile applications

The key challenge in eHealth mobile applications concerns the privacy protection of patient data which is susceptible to attacks from external attackers and from third party service providers with whom companies share data. Medical data is very prone to attacks

Table 4.10: Deriving privacy requirements from threats

From threats to privacy requirements	
eHealth Heart beat monitor threats	Privacy requirements
Linkability of patients to Heart beat monitor DB	Unlinkability of patients to database
Linkability of patients to data flows	Unlinkability of patients to data flows
Linkability of patients to Heart beat monitor processes	Unlinkability of patients to system processes
Linkability of external entities	Unlinkability of external entities connecting to systems
Identifiability of patients to Heart beat monitor DB	Pseudonymity or anonymity of patients in DB
Identifiability of patients to data flows	Pseudonymity or anonymity of patients in data streams
Identifiability of patients to Heart beat monitor processes	Pseudonymity or anonymity of patients in systems processes
Non-repudiation of patients sending data streams	Acknowledgement of patients using system processes
Non-repudiation of patients using system processes	Acknowledgement of patients using system processes
Detectability of patients connecting to DB	Unobservability of patients connecting to DB
Detectability of patients sending data streams	Unobservability of patients sending data streams
Detectability of patients using system processes	Unobservability of patients using systems processes
Information disclosure of patients connecting to DB	Confidentiality of patients connecting to DB
Information disclosure of patients sending data streams	Confidentiality of patients sending data streams
Information disclosure of patients using system processes	Confidentiality of patients using system processes
Unawareness of patients about content given in systems	User awareness of content given to systems
Non-compliance of systems with policies and regulations	Compliance with policies and regulations in systems

as it includes a lot of financial data that is embedded in the data stolen. There is a high potential of information in health data records as a lot of players are interested in accessing it ranging from medical insurance companies, employers and criminals that are interested to sell the data. Therefore, preserving privacy is very crucial and should be maintained through out the development and integration of the product life cycle. This is quite hard to sustain as their are very many players in the eHealth development Eco-system, so protecting privacy at all levels of system development with all the key stake holders involved has proven to be very challenging. As a result, the number of privacy breaches affecting eHealth applications has increased mostly at health care service providers such as hospitals and physician offices. Below we highlight some of the privacy challenges met when designing eHealth mobile communication systems.

Lack of privacy regulation and policy enforcement:

The eHealth mobile sector has not been well regulated. With various medical bodies involved, system regulation with each body varies [114] with no privacy guidelines and so application designers have been faced with a myriad of regulations to implement. In addition, the rampant development of the industry makes existent policies e.g Fair Information Practices(FIPPs) inadequate for emerging technologies. This has led to tech giants and application designers having to build applications that lack privacy protection

measures. This however, is gradually improving with efforts from tech giants e.g. Google who insist that application designers who place their applications in their app stores must ensure that they have a privacy policy for all apps in the market store.

Lack of privacy preserving mobile connection systems

eHealth mobile applications use wireless connectivity to provide remote access to patients and to connect to medical devices remotely. Wireless connectivity e.g. WiFi and Bluetooth is used to connect to terminals and body eHealth sensors which communicate with smart phones and remote servers used to store data. The wireless communication setup comes with a lot of privacy challenges and is prone to privacy and security attacks.

4.4 Empirical study 2 - Research Empirical study

This section presents the works performed in our second empirical study which involved an extensive research investigation of both eHealth and Telematics mobile applications. This study attempts to understand how mobile applications are designed and to which extent privacy protection methodologies are implemented. It investigates the implementation of privacy between the company website and the policy relating to the app. We investigate the the policies on the website versus the policies in the apps stores and those in the apps. After carrying out exploratory study 1, we found that more research had to be done in order to determine the extent at which mobile applications collect data and integrate privacy preserving mechanisms. The study involves a detailed analysis of data collection and privacy policies employed in current eHealth and Telematics applications.

4.5 TRANK based design for eHealth applications

4.5.1 eHealth case study

Within the eHealth case study, we investigate the integration of privacy in modern mobile eHealth apps. Mobile eHealth apps or digital eHealth apps are currently on the increase. They have a great potential in improving health care for patients in remote areas who cannot have a face-to-face visit with a physician. Mobile phones are used to provide online

medical platforms using eHealth apps, to assist in health monitoring and assessment of patients. This type of monitoring promises cheaper implementation costs and deployment especially as the number of doctors to patient ratio has reduced in recent years. However, the increase of eHealth applications comes with privacy concerns. The use of smartphones as a means of attending to patients remotely puts patients at the risk of their data being intercepted or landing in the wrong hands. In this case study, we investigate the highest revenue generating eHealth apps that are on the market and analyze privacy invasive capabilities in mentioned apps. We describe a taxonomy of eHealth apps which we use as a criteria in testing the apps. We further use a privacy score [77] to determine the privacy level used in the tested apps.

4.5.2 Data collection in eHealth mobile applications

In order to determine the type of data collected by current eHealth apps we analysed 30 eHealth apps. The apps we examined are shown in Table 4.11 which was retrieved at the time of this study and are subject to have changed. We initially used Google Play Store to determine the type of permissions the apps use. We subdivided eHealth apps into six categories:

- Fitness apps
- Cardiology apps
- Diabetes apps
- Weightloss apps
- Depression apps
- Physician apps

Data Collection in eHealth apps:

Data collection in mobile eHealth apps was based on the functionality of the app. By comparison, of all eHealth apps we examined, fitness apps collect the most data. Fitness apps are in addition, the most used in all the above listed categories. It was observed that

Table 4.11: eHealth apps

Apps Category					
Name	**Active Users**	**Downloads**	**Type**	**Ranking**	**Data Score**
Nike Training	252,919	10,000,000	Health and Fitness	4.6	53
Freeletics	133,224	10,000,000	Health and Fitness	4.5	22
Sworkit	109,944	5,000,000	Health and Fitness	4.6	23
Garmin Connect	245,109	10,000,000	Health and Fitness	3.9	40
Fitbit	364,990	10,000,000	Health and Fitness	3.9	41
Instant Heart rate	296,610	10,000,000	Health and Fitness	4.3	47
Pulse Point	8,148	100,000	Medical	4.5	14
Daily Cardio Workout	14,981	1,000,000	Health and Fitness	4.4	20
Nokia Health Mate	29,441	1,000,000	Health and Fitness	3.6	39
Cardiograph	200,564	10,000,000	Health and Fitness	3.8	30
Diabetes PA	1,218	50,000	Health and Fitness	4.2	38
Dexcom	38	10,000	Medical	3.5	16
Diabetes Tracker	945	100,000	Health and Fitness	3.9	19
MySugr	21,839	1,000	Medical	4.6	32
Glooko Diasend	122	10,000	Health and Fitness	2.8	30
Ideal weight	112,479	5,000,000	Health and Fitness	4.2	3
Lose it	70,373	10,000,000	Health and Fitness	4.4	34
Weight watchers	279,961	5,000,000	Health and Fitness	4.2	37
Noom coach	169,789	10,000,000	Health and Fitness	4.3	29
Calorie counter MyFitnessPal	1,900,605	50,000,000	Health and Fitness	4.6	29
Depression CBT	1400	100,000	Medical	4.2	17
7 cups of tea	14,139	500,000	Health and Fitness	4.2	23
Operation reach out	24	1,000	Health and Fitness	4.0	8
Mood Scanner	969	100,000	Entertainment	3.3	11
Fight Depression naturally	314	50,000	Health and Fitness	4.6	13
Medscape	51,272	5,000,000	Medical	4.4	15
PEPID	323	100,000	Medical	3.9	11
Uptodate	5,946	500,000	Medical	4.4	12
Aminion Doximity	1,094	50,000	Medical	4.3	17
ReadbyQxMD	1,770	100,000	Medical	4.6	8

Table 4.12: Number of apps downloaded in a category

Category	**Total No of Apps**	**>50K Downloads**	**Percentage**
Health & Fitness	80756	4004	5%
Medical	36222	1336	4%
Dating	3971	476	12%
Photography	60372	8197	14%
Role Playing	7806	2715	35%
Business	162373	3031	2%
Education	216771	12337	6%
Simulation	25615	9056	35%
Weather	9315	1337	14%
Maps & Navigation	32594	2597	8%
Music & Audio	157440	7986	5%
Social	50878	3502	7%

users who download apps, however, do not use the majority of the apps they acquire from app stores. Although the number of apps downloaded and the volume of apps increases as the number of smartphone users increase, it is estimated that less than 20% of the people who download the apps use them regularly. There are many categories of pps in Google Play Store [115], however, we derived our own categories in an attempt to differentiate them from the others and be specific in our reporting. Table 4.12 shows some of the categories in Google Play Store and the number of apps with more than 50,000 downloads. As you can observe, from the table the volume of apps in Google Play Store is not in correlation with the percentage of the apps that are frequently downloaded by users. We notice that even though educational apps have a large volume of apps in Google Play Store the percentage that is downloaded repeatedly is very low. Whereas the Simulation and Role playing apps are downloaded constantly although they are few in numbers in the store. Thus, users perception of using a pps is determinant on the functionality of the app rather than the numbers accessible in the store. It is nonetheless, important to note that data collection in the downloaded apps continues even when the app is not frequently used, and sometimes when the phone is off. Whenever users accept the permissions to the apps normally on the app downloads permission view, data collection is done to enable proper app functionality. Users are then bound to the app permissions and only with the help of the privacy policies can one choose to use the app or not. But at this stage of the app download (thats if the app has an in-app privacy policy), users can only consent to the privacy policy when most of the permissions used for e.g., device hardware data, location data, payment type data are already collected. Privacy policies tend to be long as a result, users often do not read them. Most of the privacy policies are vague and cannot be reliably understood by non-tech users. In addition, many phone screens are small and often not well lighted to enable proper readability. It follows that the majority of users do not read in-app privacy policies but constantly press the "Accept" button in order for them to continue using the app as desired. We observed that some Apps continue on accessing user information even when they are deleted from the phone. Also, phones periodically receive location based messages, information messages even when the

apps are not in use without users knowledge. Therefore, users should be informed of these activities even during app downloads to enable them opt-in or out of these services. These scenarios call for app developers to re-evaluate the way apps are developed especially by creating user awareness of which data is being collected about them. Better privacy enhancing methodologies are therefore required in an effort to achieve this goal.

4.5.3 Data extraction in eHealth mobile applications

Data extraction: The initial step of the data extraction was to download the apps from Google Play Store. We then analyzed the information given in the store about the app. Next, the privacy policy given in the store was examined and screen shots were taken in an effort to determine which data is mentioned that the app collects. Information about the data collected in the privacy policy was transcribed. After downloading the apps, accounts were created for each app and the information requested in the app was recorded. This information included;

- If the app required personal information like the name, email, address.
- A login using social networks like Facebook.
- If the app required payment information like credit card details.
- If the app had a privacy policy during log on or not.

This data enabled us to compare both privacy policies in the app and that in the store to determine any inconsistencies of personal data collected by the apps. Next we examined the privacy policies in the company websites to determine if they differed from those in the app and the store. Extracts of the data collected can be shown in Table 4.13 and Table 4.14.

Permission Use:

In addition to the collected data in both the apps and privacy policies, we recorded the permissions the apps required in the Google play store. The permissions which were required after logging in the app were further examined and compared to those in the store.

We log web browser information(e.g. browser types,... page view tallies,time spent on each page, geographic location information, page browsing information,subject browsing information and operating system information...we may collect hardware information related to the device used by the member.

Table 4.13: Snippet from 7 cups Privacy policy Browser information

...we collect personal information ...known as Personal Data and includes your name or alias or contact phone number. We also collect meta-data about the overall site, like frequency of page views...we collect information about your computer like your IP address and browser information.

Table 4.14: Snippet from 7 cups Privacy policy

4.5.4 Permissions in eHealth applications

In an attempt to understand the permissions in mobile applications, we need to understand what type of permissions exist and their purpose. Most users cant distinguish permissions that are required for app functionality from permissions that are used to monitor user movement and their activities.

What are permissions used for? Permission requests are used to protect sensitive data on the gadgets. Users can choose which permissions the app has access to. However some permissions are mandatory for the basic functionality of an app. Take for example, a telematics insurance app requires the position information (geo-location using GPS sensors on the phone) for the app to know a drivers' exact location in order to measure the distance traveled and calculate the premium based on these details. Similarly, to download apps a connection to the Internet is unavoidable, so the majority of apps require the "full network access" permission.

Myriad of app permissions required: The plethora of app permissions in the app store is overwhelming for a non-tech user. Much as most of the permissions are listed in Google Play Store, its really hard to comprehend what these apps mean for any lay person. The actual number of permissions an app requires does not depict which information is going to be used and collected from the phone and why its needed. Various app permissions

Table 4.15: Example of Permissions extracted

Permissions	
read your own contact card	heart rate monitors
add or remove accounts	read Calendar events plus confidential information
read sensitive log data	receive data from internet
find accounts on the device	view network connections
read your own contact card	pair with Bluetooth devices
read your contacts	access Bluetooth settings
modify your contacts	control Near-Field Communication
read phone status and identity	read Google service configuration
access USB storage file system	toggle sync on and off
directly call phone numbers	download files without notification
read call log	read sync statistics and settings
write call log	full network access
access USB storage file system	control flashlight
read the contents of your USB storage	control media playback and metadata access
modify or delete the contents of your USB storage	smart card services permissions label
read the contents of your USB storage	send sticky broadcast
modify or delete the contents of your USB storage	change network connectivity
read phone status and identity	connect and disconnect form WiFi
approximate location (network-based)	change your audio settings
precise location (GPS and network-based)	run at startup
access extra location provider commands	draw over other apps
receive text messages(sms)	control vibration
read your text messages(SMS or MMS)	prevent device from sleeping
receive text messages(mms)	modify system settings
send sms messages	Google play licence check
read Calendar events plus confidential information	reorder running apps
take pictures and videos	prevent device from sleeping
record audio	close other apps
install shortcuts	uninstall shortcuts
manage document storage	create accounts and set passwords
view WiFi connections	use accounts on the device

require a lot of PII and a collection require only information about hardware components. Some of the permissions recorded were given as follows:

1. Identity: find accounts on the device, read your own contact card, add or remove accounts.

2. Phone: read phone status and identity, access USB storage file system, directly call hone numbers.

3. Location data: approximate location (network based), precise location (GPS and network based) and access extra location provider commands.

Table 4.16: Nike Training Permissions

	Nike Training		
Permissions		*Permissions*	
Identity		Other	
	find accounts on the device		download files without notification
	add or remove accounts		read sync statistics
Contacts			receive data from Internet
	find accounts on the device		view network connections
	read your contacts		create accounts and set passwords
Location			full network access
	approximate location (network-based)		control Near Field Communication
	precise location (GPS and network-based)		read sync settings
Phone			run at startup
	read phone status and identity		use accounts on the device
Photos / Media / Files			control vibration
	read the contents of your USB storage		prevent device from sleeping
	modify or delete the contents of your USB storage		toggle sync on and off
Storage			
	read the contents of your USB storage		
	modify or delete the contents of your USB storage		
Camera			
	take pictures and videos		
Wi-Fi connection information			
	view Wi-Fi connections		
Device ID & call information			
	read phone status and identity		

Table 4.17: Depression CBT Permissions

	DepressionCBT		
Permissions		*Permissions*	
Phone		Other	
	read phone status and identity		receive data from Internet
Photos / Media / Files			view network connections
	read the contents of your USB storage		full network access
	modify or delete the contents of your USB storage		prevent device from sleeping
Storage			
	read the contents of your USB storage		
	modify or delete the contents of your USB storage		
Device ID & call information			
	read phone status and identity		

We recorded these permissions for each app and observed that some apps requested for permissions which were not used for application functionality. There is a tendency of collecting more data than is required for system functionality but rather gather more information for advertising and system analytics. Nike Training for example, required more permissions than Depression CBT as shown in Table 4.16 and Table 4.17. Mysugr required few permissions as compared to Nike Training as depicted in Table 4.18 although it provides a lot of features. Therefore, permissions requested varied from app to app based on not only the services given but also on how much advertising and systems analytics is done for a particular app and company.

In an attempt to investigate what type of data is collected in the apps, we studied the

Table 4.18: MySugr Permissions

	MySugr		
Permissions		*Permissions*	
Identity		Other	
	find accounts on the device		receive data from Internet
Contacts			view network connections
	find accounts on the device		pair with Bluetooth devices
Location			access Bluetooth settings
	approximate location (network-based)		full network access
	precise location (GPS and network-based)		run at startup
Photos / Media / Files			control vibration
	read the contents of your USB storage		prevent device from sleeping
	modify or delete the contents of your USB storage		
Storage			
	read the contents of your USB storage		
	modify or delete the contents of your USB storage		
Camera			
	take pictures and videos		
Other			

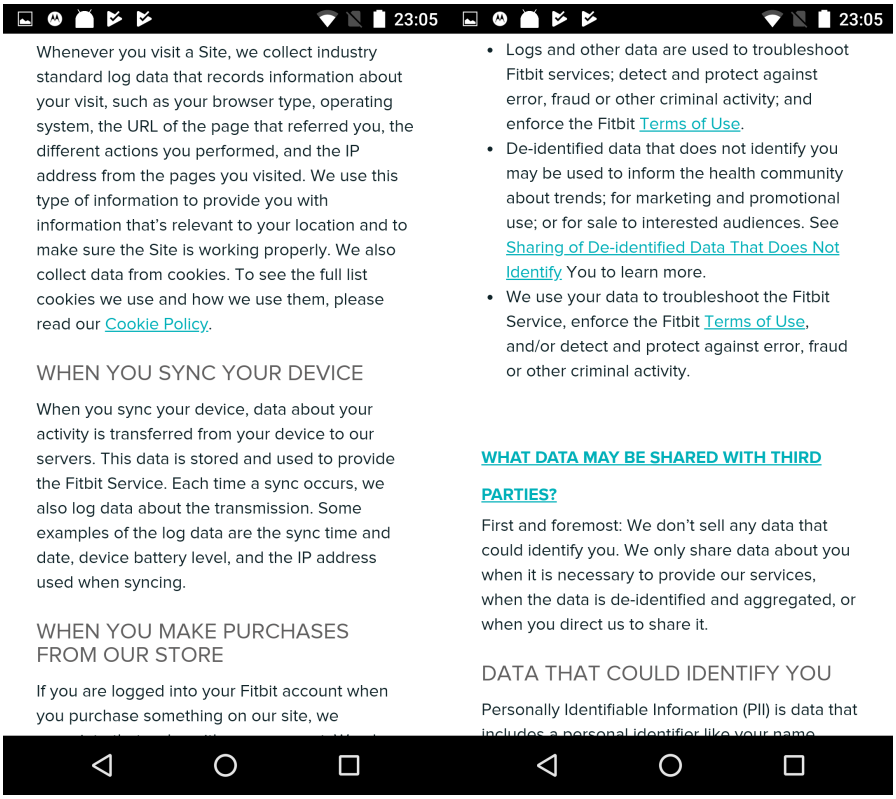


Figure 4.7: Fitbit privacy policy - sharing data with third parties.

privacy policies to find out what is included in the policies. We further recorded details of the privacy policy of how data is shared to third parties for the Fitbit app as seen in Figure 4.7 which states that "*First and foremost we don't sell any data that could identify*

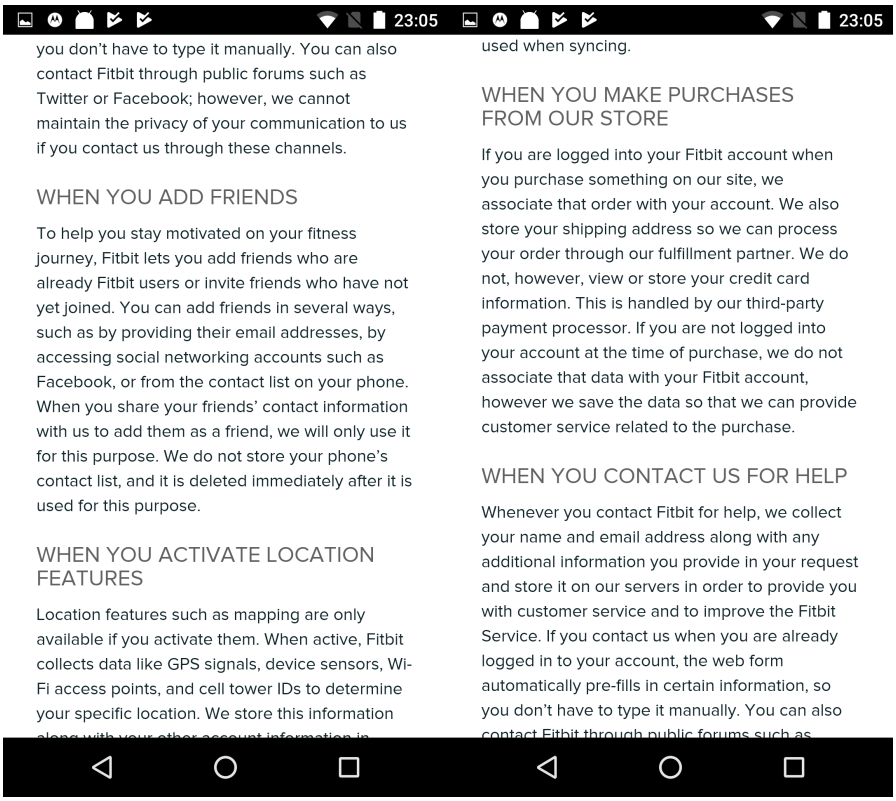


Figure 4.8: Fitbit privacy policy - of how data is shared with friends and location details.

you. We only share data about you when it is necessary to provide our services, when the data is de-identified and aggregated, or when you direct us to share it." This statement is not clear on what and when data is shared. We observed that most privacy policies were vague and hard to understand by a normal user. Details of the type of data shared and how its protected are not given. This trend was observed in most of the privacy policies we analysed.

In Figure 4.8 information about how data is shared with friends using the same app is given as follows:"*To help you stay motivated on your fitness journey, Fitbit lets you add friends who are already Fitbit users or invite friends who have not yet joined. You can add friends in several ways such as by providing their email addresses, by accessing social networking accounts such as Facebook or from the contact list on your phone.*" These statements involve data sharing to both friends in the contact list and in Online Social Networks without user consent. eHealth apps have further advanced to use genetic data to offer services based on DNA as is with the Lose It App which used genetic data to

Figure 4.9: EmbodyDNA Kit

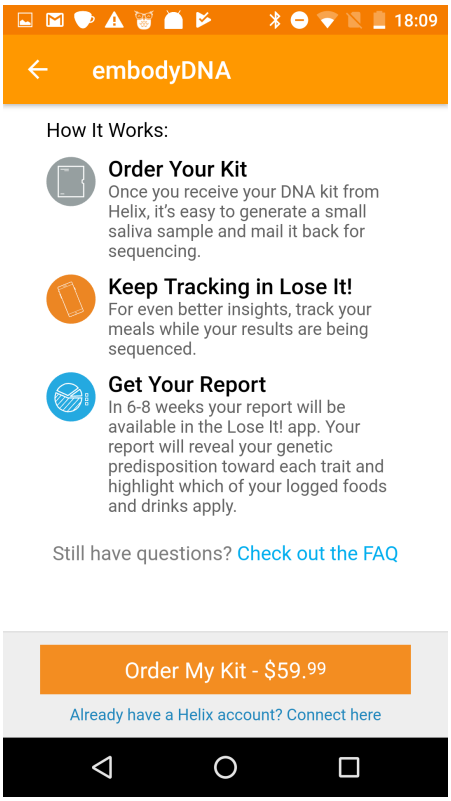


Figure 4.10: DNA Report

determine if customers can lose weight depending on the genetic DNA results.

Figure 4.9 included details of how the genetic traits may determine weight loss as follows; "*Learn your unique genetic traits and how they could affect your weight loss, nutrition, fitness, and food sensitives*" and Figure 4.10 states that; "*In 6-8 weeks your report will be available in the Lose it! app. Your report will reveal your genetic predisposition toward each trait and highlight which of your logged foods and drinks apply.*" These details are processed without user consent and the fact that genetic data is saved in the app makes it vulnerable to attackers who may steal customers genetic data and compromise user privacy.

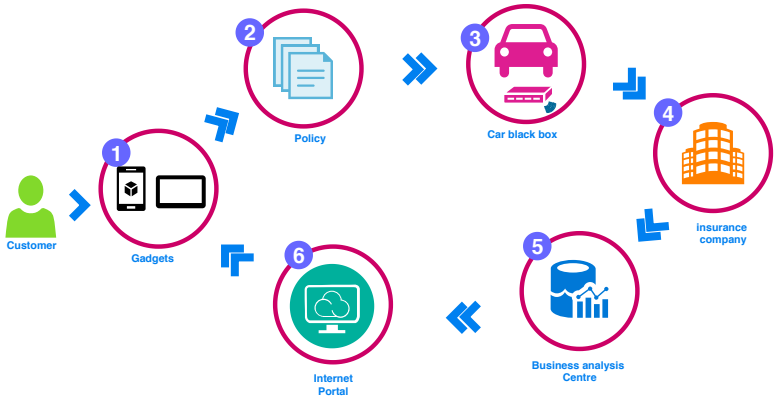


Figure 4.11: How Telematics insurance works

4.6 TRANK based design for Telematics applications

4.6.1 Telematics insurance case study

In our second case study we focus on the emerging new technologies in Vehicle-to-Everything (V2X). The V2X case study we used is that of a telematics insurance application. Figure 4.11 illustrates how telematics insurance works. In a telematics insurance application users subscribe to the application and in return are charged less fees if they drive less frequently and carefully. The main objective of telematics insurance is to provide cheaper costs of car insurance to young drivers between the ages of 17 to 25 years. Compared to conventional insurance policies that are billed based on the Pay-Per-Year (PPY) model, telematics insurance is billed according to the number of miles driven; Pay-Per-Mile (PPM). It is further billed depending on what time you drive and how you drive; Pay-AS-You-Drive (PAYD). This payment model makes it cheaper and attractive for young drivers who would like to reduce initial insurance costs. However, telematics insurance has faced a lot of criticism regarding its 24 hour surveillance and tracking of its users. This has been seen as privacy invasive as users are not aware of how the data is used especially if its sold to third parties.

4.6.2 Data collection in telematics mobile applications

We analysed 10 telematics apps and recorded the type of data they collect. We used Goople Play Store to download the apps checking the type of permissions the apps requires

Table 4.19: Telematics Insurance apps

Apps Category					
Name	**Active Users**	**Downloads**	**Type**	**Ranking**	**Data Score**
Aviva drive	8,687	100,000	Finance	4.0	31
Tomtom curfer	153	10,000	Auto and Vehicles	3.0	26
Marmalade	7	1000	Finance	2.6	26
Smart Miles	366	10,000	Maps and Navigation	3.9	4
Admiral Insurance	28	5000	Business	4.3	12
RAC Telematics	71	10,000	Business	2.1	20
Smart Wheels	584	50,000	Auto and Vehicles	1.4	5
Drive Smart Insurance	1,091	100,000	Maps and Navigation	3,9	28
SEAT Telematics	26	500	Lifestyle	1.0	15
Volkswagen Telematics	66	1000	Lifestyle	1.0	22

on installing it. We further analyzed details of the privacy policies presented on company websites and in-app privacy policies. Telematics insurance technology in most of the apps analyzed is similar in comparison to the different types of the eHealth apps we analyzed, which offer diverse features to customers. Therefore, the technical type of data collected was much the same. However, the amount of data required by the apps varied immensely.

4.6.3 Data extraction in telematics mobile applications

The first step of the data extraction was to download the apps from Google Play Store. Figure 4.19 shows details of the apps we analyzed from Google Play store. We recorded the information given in the store about the app. Next, the privacy policy given in the store was examined and screen shots were taken in an effort to determine which data is mentioned that the app collects. Information about the data collected in the privacy policy was transcribed.

After downloading the apps as shown in Table 4.19, accounts were created for each app and the information requested in the app was recorded. The data collected included;

- Whether the app collected PII information i.e, name, email address, telephone numbers
- Data shared to Online Social Networks
- Whether the app collected financial information e.g employer, debit card details

- Whether the app had no discrepancies in the privacy policies during sign up ad the privacy policies given on the company websites.

Similar to the data collection in eHealth apps, we used Google Play Store to determine the permissions the telematic apps have access to. We compared the privacy policies presented in Google Play Store to the privacy policies presented on the company websites. We installed the telematics apps investigated on a Motorola XT1092 phone running Android version 8.0. Our dataset consisted of the most popular telematics apps. We collected data from 10 telematics insurance companies namely, Aviva Drive [116], Tomtom Curfer [117], Marmalade [96], Hastings Direct - Smart Miles [118], Flo driving insights [119], Smart Wheels [120], Insure the box [98], Drive smart Insurance [99], Seat Telematics [121] and Volkswagen telematics [122]. We observed that most of the apps required users to sign in for insurance policies. We therefore, used the privacy policies presented in the company sites for our analysis. We transcribed the information given in the privacy policies to determine which data is collected by the apps. Our aim was to determine the following type of data collected;

- Which permissions does the app have access to?
- Are the permissions requested required for app functionality?
- Does the app have a privacy policy in Google Play Store?
- Does the app have an in-app privacy policy?
- Does the app have a privacy policy on the companies website?
- Does the privacy policy contain the details of the data that it collects?
- Does the privacy policy contain the contact of the developer?

We analysed the information given in the store about the apps both in the companies websites and the apps. This was done in two phases; (1) First, we examined the permissions required while downloading the app and (2) the second phase was to examine the privacy policies.

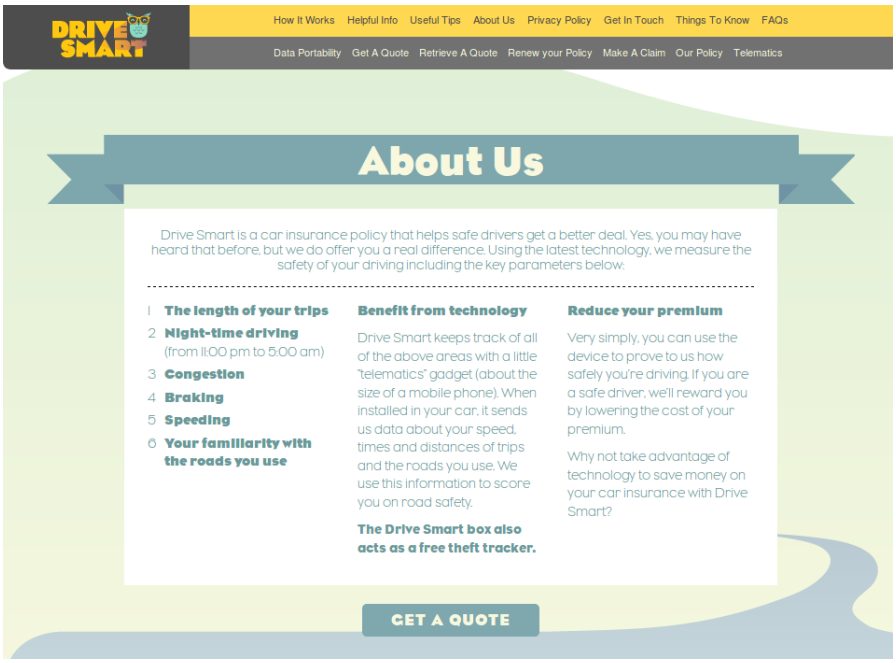


Figure 4.12: Data collection in Drive smart telematics insurance.

As illustrated in Figure 4.12 we have extracted data given in websites that informs users of what the companies collect. We found that Drive smart explicitly informs users of which data they collect. This includes the length of the trips taken to determine if the driver travels long distances on the highway. This information can be used to predict if the driver is prone to making accidents, thus increasing the price they have to pay for the insurance. Drive smart further collects data about driving at night between 11:00 pm and 5:00 pm. This information is also necessary to determine if the driver may cause accidents at night. Drive smart further collects data about customers driving in congested areas, it collects data about how often they brake, or speed up and finally how familiar they are, with the roads they drive on. This is important as drivers tend to make mistakes in new environments because they are not used to the roads and thus increasing the probability of causing accidents. We transcribed this data to enable us determine if there were no discrepancies between the privacy policies given on the websites and in-app privacy policies.

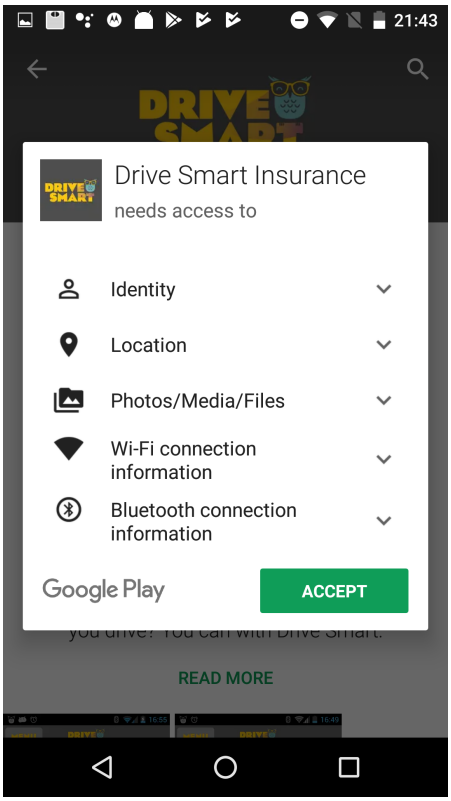


Figure 4.13: DriveSmart permissions

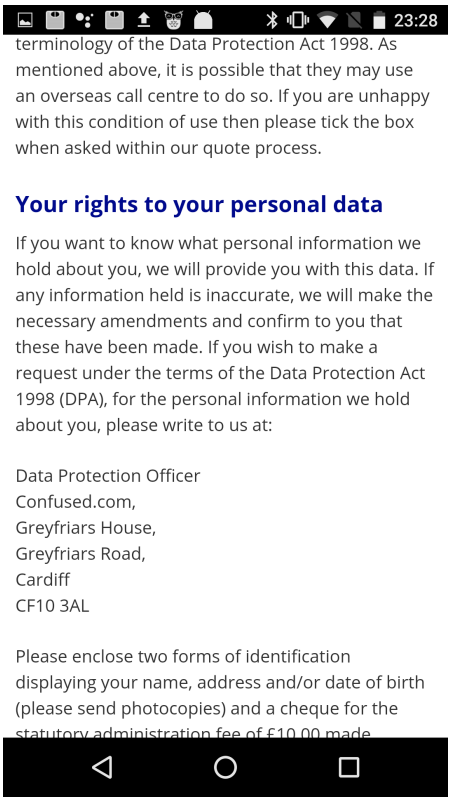


Figure 4.14: User rights.

4.6.4 Permissions in Telematics applications

We manually recorded the permissions the apps required in the Google play store. The permissions required after logging in the app were further examined and compared to those in the store. Figure 4.13 shows the in-app permissions for the Drive smart telematics application and Figure 4.14 shows the rights to personal data.

On installing the application we made screenshots of the application similar to what we did for eHealth applications. We examined all permissions of the application as presented in the Google Play Store and within the app. The permissions are as follows;

1. Storage
    - modify or delete the contents of your USB storage
    - read the contents of your USB storage
2. Location

Table 4.20: Marmalade Telematics insurance Permissions

	Marmalade		
Permissions		*Permissions*	
Device and app history	read sensitive log data	Other	receive data from Internet
Contacts	find accounts on the device		pair with blue tooth devices
	read your contacts		access blue tooth settings
Location	approximate location (network-based)		full network access
	precise location (GPS and network-based)		manage document storage
Phone	read phone status and identity		
Photos / Media / Files	read the contents of your USB storage		
	modify or delete the contents of your USB storage		
Storage	read the contents of your USB storage		
	modify or delete the contents of your USB storage		
Camera	take pictures and videos		
Wi-Fi connection information	view Wi-Fi connections		
Device ID & call information	read phone status and identity		

- precise location (GPS and network-based)
- approximate location (network-based)

3. Wi-Fi connection information
   - view Wi-Fi connections
4. Phone
   - read phone status and identity
   - directly call phone numbers
5. Photos/Media/Files
   - modify or delete the contents of your USB storage
   - read the contents of your USB storage
6. Device ID and call information
   - read phone status and identity
7. Other

- download files without notification
- receive data from Internet
- view network connections
- pair with Bluetooth devices
- full network access
- access Bluetooth settings
- control vibration
- prevent device from sleeping
- run at startup

We transcribed all telematics permissions as seen in Table 4.20 for the Marmalade telematics insurance app. We opted to record all the permissions as given in the websites and compared them with those in the apps to determine any asymmetries. For each app we checked for any differences between the permissions in the Google Play Store websites and those reported in the apps. Permissions play a major role in determining which data the app has access to. For example, we observe that although most of the permissions requested are used for functionality such as the permission request *access to Wi-Fi connections* is used to access the internet, and *access to the camera* is used to enable taking photos, some of the permissions used are entirely for marketing and advertising. A Global Privacy Enforcement(GPEN) survey [123], reported that most of the permissions requested are not required. This means that app designers explicitly ask for more data that is not mandatory for app functionality. This results into immense amounts of data being collected without user influence and empowerment. Although some apps include details of having access to the data collected within the app, this comes at a fee as depicted in Figure 4.14 and so customers are hesitant to ask for their data. According to the GPEN survey [123] 32% of apps had access to location data and the number is increasing. This means that customers are constantly being monitored wherever they go making it very lucrative for data analysis used for determining user

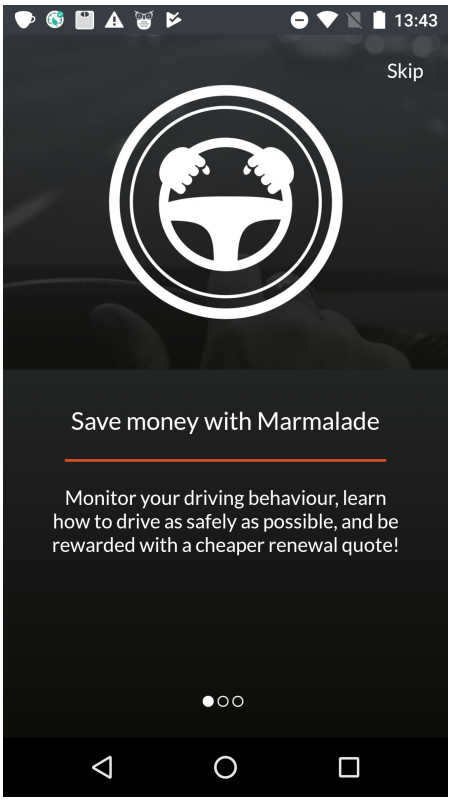


Figure 4.15: How the app is used.

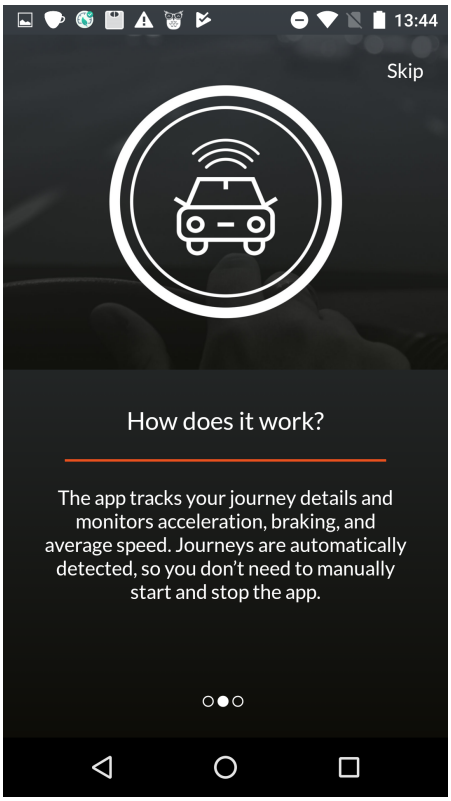


Figure 4.16: How the app works.

trends. In the telematics insurance domain location information is required to determine the cost of the insurance policy based on the customers driving.

In-app privacy policies in Telematics applications

According to the new applicable laws e.g GDPR, apps are required to have a privacy policy within the app that explicitly informs users of how user data especially PII is handled. Google Play Store demands that all apps have a link to the privacy policy within the store. We observed that the majority of the apps had a privacy policy on the company sites as well as on Google Play store but lacked one within the app. Apps are required to give information about the data they collect and about what the app does as illustrated in Figure 4.15 and Figure 4.16. However, the majority of the apps did not have this information within the app but had a privacy policy or a link to the privacy policy in the app. We observed that generally companies provide in app privacy policies however, these policies are too long for some of the devices used. This means that users do nor fully read the policies they are provided.

For instance, Figure 4.17 is part of a long privacy policy with phrases that are hard to comprehend for layman users. Similar to the privacy policies on Google play, it illustrates how the data collected is used which states as follows:

"*We may use your data to review and analyse market trends and to track sales data, including but not limited to data provided as part of your quotation, information on how you use our website, pages you have viewed and which websites you used previously to visit ours. We may also pass this data to third parties for analytical purposes. Where you have come to our website by clicking on a price comparison service button or another service button on another site, we will supply information about the quotations you have obtained and any policies you obtain to that site....*"

This statement justifies that most mobile apps share information to third party service providers for analytics and marketing. Although the legal basis for data collection enacted by the GDPR regulations and the EU law demands that mobile app developers include for which specific purposes the data collected is for, we observed that most of the apps did not present this information.

Not only is data shared to third parties within the UK but also in other countries outside the UK as illustrated in the privacy policy shown in Figure 4.18 which states that: " *All personal information you provide will be held in the strictest confidence and only used for the purpose of providing our service, subject to certain exclusions as described within this Privacy Policy and the Terms and Conditions. From time to time there may be a requirement to process your personal data in other countries out side the European Economic Area (EEA) where data protection safeguards differ from those of the UK. Where this is necessary we will ensure that your data is kept securely and only processed in accordance with the Data Protection Act. Some of our service providers may also process your data in other countries and we require all your data is kept securely at all time and the same UK standards are met.*" This statement clearly shows that users have no control over their data when companies have shared it outside the EEA. Cross-boarder data transfers are regulated by the GDPR in article 45[124] and 46 [125]. Article 45 section 1 states that: "*A transfer of personal data to a third country or an international*

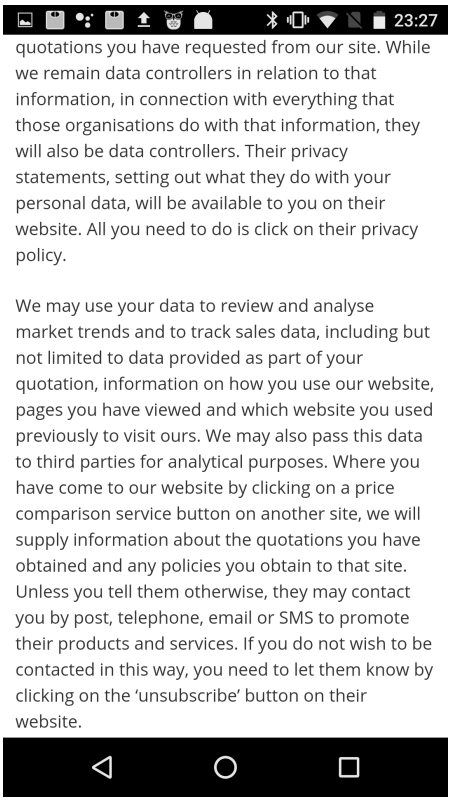


Figure 4.17: In app privacy policy showing how data is used.

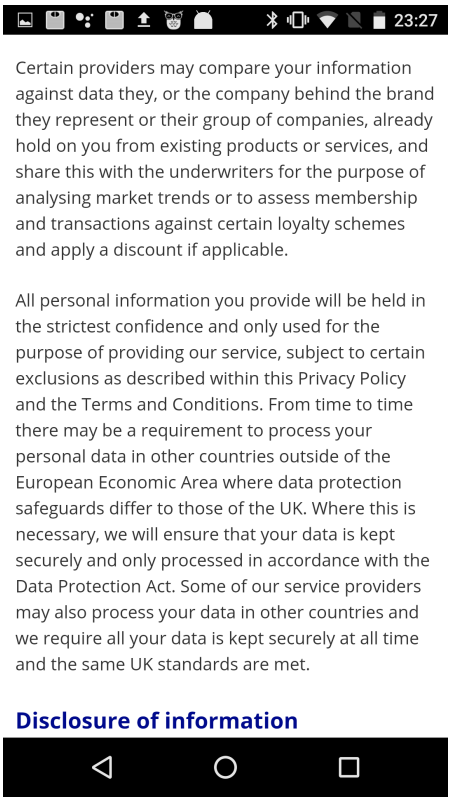


Figure 4.18: In app privacy policy data sharing to third-parties

organization may take place where the Commission has decided that the third country, a territory or one or more specified sectors within that third country, or the international organization in question ensures an adequate level of protection. Such a transfer shall not require any specific authorization." This clearly states that transferred data has to be protected. This information however was lacking in most of the privacy policies we analyzed.

4.7 Privacy metrics used for Data Analysis

To compute the data transcribed in the studies above we had to choose a privacy metric that best suits the type of data collected from privacy policies, the functionality features provided by the apps and the amount of data collected.

The privacy metrics used in this work are based on the analysis in [77]. The analysis

relies on the amount of data collected, the privacy guidelines used in the privacy policies and the privacy controls employed by the apps. These metrics are means to quantify the usefulness of our design for the TRANK framework. The privacy score (PS) is a measure of how much privacy is implemented in a particular app. This value is obtained by computing the average of the privacy control score (PCS), the privacy policy score (PPS) and the data collection score (DCS) as follows:

$$PS = \frac{(PCS + PPS + DCS)}{3} \tag{4.1}$$

$$PCS_2 = \frac{PCS_1 - min(PCS_1)}{max(PCS_1) - min(PCS_1)} \tag{4.2}$$

$$PPS_2 = \frac{PPS_1 - min(PPS_1)}{max(PPS_1) - min(PPS_1)} \tag{4.3}$$

$$DCS_2 = 1 - \frac{DCS_1 - min(DCS_1)}{max(DCS_1) - min(DCS_1)} \tag{4.4}$$

The three subscores are further defined as follows;

Privacy Control Subscore: is derived from the features integrated in the app and the company site to ensure that their privacy is protected. An example of such features include the use of encryption measures e.g Transport Layer Security (TLS) or how the privacy configuration is set in the app. This entity was given a value of 1 point.

Data Collection Subscore: is calculated using the number of data items that are collected in the app. The data items are further computed from the details given in the privacy policies. This is to ensure that the data items that are mentioned in the privacy policies and not clearly seen during app download or when the app is running are not missed out. This entity was given a scale of 1 for each data entity and varied from 1 to 3 based on the entity it represented. For example, we allocated a value of 1 to data entities e.g name, gender and a value of 3 for the use of social networks in an app.

Privacy Policy Subscore:

The Privacy Policy Subscore is computed based on if the app or the company website has implemented a privacy policy and the details included in the app. The details were allocated a value of 1 e.g. based on if the privacy policy existed in the app, within the app and within the company site. Other factors were put in consideration e.g. if terms and conditions existed, if personal information can be deleted.

The derived values were further normalised to give each score equal weight for our analysis. Data collection scores (DCS) are reversed as a higher data collection score leads to a decrease in data privacy.

Initially, Google play store was used to list the permissions that apps are able to access. A subsequent step was to award points to the permissions that we deemed privacy intrusive based on the app functionalities. For instance, the point range from -0.5 to 3 was used to share data with third parties and to connect to social platforms respectively. Next, PCS values were analysed which involved among others checking for data security features like the use of Transport Layer Security (TLS) during login. PPS values were determined based on privacy policies which were rated using accessibility, use of privacy policies during sign on and availability of contact details. Our data set consisted of 30 eHealth apps. Each app was downloaded from Google play store onto a Motorola XT1092 phone running Android version 8.0. Then privacy policies in Google play store as well as the company sites were analysed. Further, the functionality the app provided and the privacy controls that were implemented in the app were analysed. The results were digitally recorded and transcribed in three categories, one containing the results for the PPS, PCS and DCS.

4.8 Summary

This section provides the contributions of this thesis in which we analyse privacy preserving methodologies employed in current eHealth and Telematics applications. We initially analyse the eHealth domain. In the mobile application selection and analysis section, we categorized eHealth apps in six independent categories, namely; Fitness apps, Cardiology

apps, Diabetes apps, Weightloss apps, Depression apps and Physician apps. We focused on these apps as they are the leading apps in use with high revenues and due to their highly evolving nature. In particular, we focused on 30 eHealth apps and 10 telematics apps, in which we examined the permissions reported on the company websites, permissions in Google Play and in-app permissions. We used this approach to investigate if there are any asymmetries and inconsistencies in the permissions provided to customers. We investigated which permissions are used and what they are used for within the app and in Google Play Store and if the app permissions requested for, are required for app functionality. Next, we analysed privacy requirements analysis in eHealth apps with a focus on data extraction in eHealth applications, in particular the permissions required in eHealth apps and in-app privacy policies used. Finally, we investigated the telematics insurance domain starting with the privacy requirements analysis, followed by data extraction, permissions and in-app privacy policies. In general, this section gives a comprehensive investigative study of privacy-by-design methods used in current mobile applications. In the following chapter, we introduce TRANK a privacy trade-off analysis framework which aims at addressing the gaps discussed in chapter 1, 2 and 3 by enhancing the privacy-by-design approaches used in mobile applications.

Chapter 5

TRANK: A privacy-aware framework for designing future mobile applications

5.1 TRANK framework overview

This chapter presents a privacy Trade-off Analysis Framework (TRANK). TRANK is a design framework for designing privacy aware mobile applications. Designing of privacy-aware mobile applications is crucial in order to protect user Personally Identifiable Information (PII) and to reduce data leakages. In practice, mobile apps collect a lot of data required for proper system functionality ranging from personal data e.g. name and address to billing information e.g credit card details. Privacy data protection reduces the damages caused when such systems are compromised. Today data leakages in mobile applications like social networking services (e.g. Facebook) and web service providers (e.g. Yahoo) are reported frequently and the cost of the damage is increasing. As such, the results in chapter 1 show that existing mobile applications are prone to massive data leakages due to the large amounts of data they collect and their inability to protect this data and, therefore, require novel privacy protecting methodologies.

Privacy awareness has become a major concern in recent years as companies continue

to collect data without users consent. We have seen a continuous monitoring of users through third-party companies which are outsourced as discussed in chapter 1. In recent years, privacy concerns have increased about the use of personal data by data controllers and processing companies. User's demographic location is used to profile users as depicted by the Cambridge Analytica data scandal [20]. In addition, we have seen a massive growth of mobile applications which collect large amounts of personal data. This massive growth raises privacy concerns both from users and legislators especially on how personal data is used. The EU General Data protection regulation (GDPR) has mandated that all companies must inform users of what type of data is collected. Much as companies have attempted to address privacy directives by using privacy policies, despite these efforts they continue to collect data without informing users.

These challenges call for a consistent methodology to design privacy-preserving mobile applications. Mobile apps have increasingly been deployed by companies to aid in, among others, the management of business efficiency, ease in accessing information, simplifying communication and the provision of user-friendly applications. However, there is a significant gap in having common practices for designing and implementing privacy-preserving methods. TRANK, therefore, aids mobile application developers and privacy engineers in designing privacy-aware applications from the initial stages of system development most importantly putting into consideration the trade-offs between privacy, system functionality and performance. We propose TRANK, to improve the development of privacy-aware apps and report on our empirical findings derived from our study of current mobile applications in the previous section. Our analysis conducted on real eHealth and V2X telematics case studies reveals that more than 50% of current apps accumulate a lot of data that is not required for app functionality without the users' knowledge. Our proposed framework TRANK thus is a contribution towards privacy preservation in current mobile applications.

Privacy integration in mobile apps: In practice, mobile apps have to consolidate privacy requirements by providing privacy policies and open privacy practices. However, a major privacy concern that arises is whether current apps depict what is in the privacy

policies they provide to their users? In an effort to close this gap, we examined the data the apps collect at the app level and that mentioned in the privacy policies and company websites. The data examined involved for example personal identification data e.g. name, email, address, date of birth; billing information e.g. billing address, credit card numbers; location based data e.g GPS location data to determine any asymmetries involved. The major aim of determining these differences is to reduce the massive data apps collect without users consent that is not necessarily required for app functionality.

This chapter aims at fulfilling this call by proposing a privacy aware trade-off analysis framework which aims at leveraging app functionality and privacy aware requirements, to aid application designers in developing more privacy aware mobile applications without minimizing app functionality.

The major contributions of this work are three fold:

1. An empirical study of 30 eHealth apps and 10 telematics apps to determine data collection inconsistencies of mobile applications at the app level compared to the data mentioned in the companies privacy policies.

2. A privacy trade-off analysis framework(TRANK) which evaluates, identifies and minimises data collection inconsistencies of data collected at the app level to that mentioned in the privacy policies and company websites.

3. A statistical evaluation of the privacy aware analysis framework (TRANK) which indicates a significant reduction of data collection in examined apps thus improving data privacy.

Given the aforementioned limitations in chapter 1 and 2, there is a significant gap in the the way apps are designed and how users data is stored and managed. The above incidences highlight the need for the design of less privacy invasive apps while considering the impact of privacy trade-offs. Proposing TRANK, therefore, serves as a stepping stone in addressing these challenges.

Privacy trade-offs in application development: Previous research has been done to tackle trade-offs in relation to security, utility, functionality and performance.

Privacy trade-offs play a great role in designing mobile applications because for some applications to function as they are required, e.g in navigation apps, location data must be used. This means that in such instances privacy has to be sacrificed. Next we present previous works on trade-offs we reviewed as a basis of our research.

Liu et al [126] introduced a model to determine the trade-off between functionality and user privacy preferences. This model was aimed at performing personalized App recommendation using three levels of privacy information. While this focuses on personalized app recommendations based on the trade-off between app functionality and user privacy preferences, the focus of our framework is on reducing data collection in mobile applications while putting into consideration the trade-off between privacy and app functionality.

Patil et al [127] conducted a survey to determine the perception of security and privacy during surveillance on the Metro in Europe. This study aimed at finding the trade-off between privacy and security in surveillance-oriented security technologies and to determine if they were privacy intriguing. This approach however defers from our research in that we aim at improving privacy preservation through limiting data collection in mobile applications. We focus on the trade-off between app functionality and privacy preserving mechanisms that are used in app development and design.

Xi et al [128] introduced a class of privacy definitions called Blowfish which entails policies used to determine the trade-off between privacy and utility. In their classification, they use kmeans clustering and cumulative histograms to tune utility in relation to privacy specifications. Policies are used to distinguish information in the data base that is to be protected against attackers thus enhancing utility. While this approach is similar to ours in the sense of considering privacy trade-offs in application development it differs to ours in terms of objectives. Our work focuses on privacy trade-offs in terms of app functionality in mobile applications and not privacy trade-offs in relation to utility in databases.

Enck et al [129] introduced a tracking system for monitoring privacy on smart phones named Taintroid. Tantroid was used to study 30 third party applications from the Android

Market place. Their study included an analysis of the Android execution environment to determine which third party applications misused users' private information. This work closely relates to ours in the sense that it was performed using Android applications to determine which third parties misuse data collected. However, it defers to ours in objective that a solution in terms of how the data collection is minimised and integrated with privacy management is not given.

In this work, we address challenges of privacy trade-offs by investigating privacy policies, privacy control measures and data collection to get an insight in the privacy implementations employed in current apps. The main purpose of our study is to provide a framework for designing privacy aware apps using the eHealth and V2X telematics domains as case studies.

Why do we need a Trade-off framework?

There has been a multitude of mobile applications developed and uploaded on to the Android market known as Google Play store. However, these mobile applications come with privacy related challenges. Android applications may reveal private data to both third-party applications or app developers through permissions. Android apps use permissions to implement security and privacy. Much as the permissions are well structured in restricting the apps not to access user's data, some apps leak personal data to third-party service providers [21]. Furthermore, users are not aware of how this data is used by app designers and third party applications. Users have to authorize third party service providers before their personal data has to be shared as envisaged by the EU 2016/679 General Data Protection Regulation (GDPR). Such privacy preserving measures have been enacted by the EU legislative bodies. However, users personal private data has continuously been shared and bought outside the devices by collecting companies and third party service providers [22]. The element of trade-offs arises when users have to choose between trading their privacy for better functionality or giving up app functionality in order to attain privacy. Therefore, there is a need for a solution that takes into consideration the trade-off between privacy, functionality and performance. For these reasons we found it necessary to develop the privacy aware trade-off analysis

framework.

Domain selection: We analyze the eHealth and the Vehicle-to-Everything (V2X) telematics domains, to determine if the data collected by mobile applications at the app level is consistent to that mentioned in the privacy policies. Further investigations to find out if the data collected is required for app functionality were performed. We put emphasis on these two domains because of the highly sensitive personal data they collect and the privacy regulations they must use for app development. Next, we highlight on eHealth and V2X privacy concerns reported with respect to privacy violations in current mobile apps on the market.

eHealth: The emergence of eHealth apps or mobile health apps has created new revenue generating opportunities for medical companies in fields like diabetic management, patients analysis and disease management to enable patients monitor their health status continuously. These apps are a big relief to customers and offer very good services towards health care management. However, The major concern though is that many of these apps do not go through stringent traditional quality controls and scrutiny from medical organizations, although they do collect a lot of private and personal sensitive data. Users are not explicitly told of the data and the implications of this data being collected. A good example of such massive data collection has been observed in fitness apps and weight loss apps. We therefore explored, this domain using TRANK in an effort to determine how privacy preservation can be improved.

V2X applications:

In V2X applications, the major privacy concern is vehicle tracking through identifying location information. This requires anonymity and unidentifiability of vehicles in the network through pseudonym use to avoid inferring personal information from drivers [130]. Cyber attacks on connected vehicles and autonomous vehicles are of major concern as well, especially, if hackers in the vehicular network are able to compromise drivers privacy [131]. The V2X applications domain used in this research, however, is that of telematics insurance. Telematics insurance applications are on demand as they provide cheaper alternative insurance policies compared to the traditional car insurance policies. Privacy

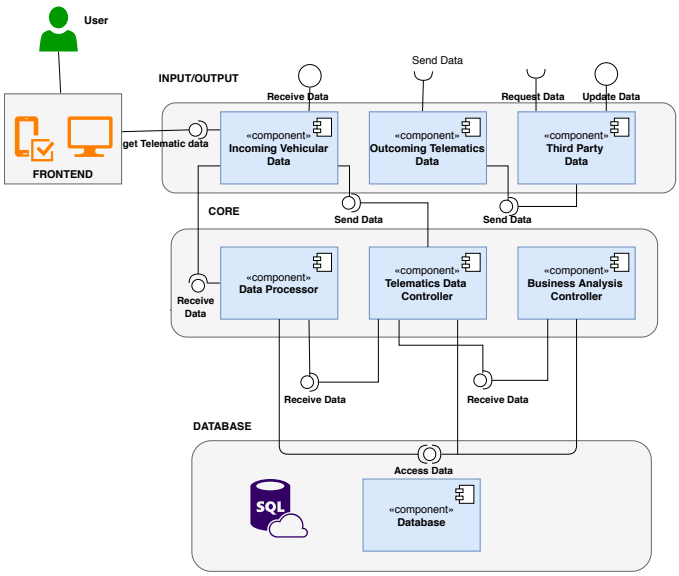


Figure 5.1: Telematics components

concerns have been raised as users are of the opinion that telematics insurance companies are constantly tracking and monitoring them. Additionally, drivers have concerns about their data being sold to third-party service providers without their consent [132]. Modern vehicular applications are composed of in built sensors which are capable of tracking our locations, social behavior and monitoring of our day to day activities. The collected data is often used to infer our personal and social movements thereby generating a lot of privacy concerns. An overview of the V2X application components and how they are interconnected is illustrated in Figure 5.1. In the following, we present the Privacy Aware trade-off Analysis framework (TRANK) to aid in solving the aforementioned privacy invasion challenges. TRANK further aims to enhance privacy preservation methodologies in current mobile applications.

5.1.1 TRANK Framework

In this section, we present TRANK, a modeling framework for designing privacy-aware applications. TRANK is motivated by the existence of a large diversity of privacy requirements when designing apps originating from different domains. The major challenge is how to design a framework which suits different domains. The framework should provide a meaningful solution to designing privacy-aware applications putting into consideration

the trade-offs between privacy and functionality.

Step 1: The initial step of the framework defines privacy, functional and performance goals of the planned system.

The major objective of our framework is to determine which goals are most important and how best they can interplay with the other goals under consideration. We emphasize Privacy-by-Design (PbD) principles [133] with a strong consideration of privacy goals especially during the initial stages of the app development life cycle as shown in the process flow in Figure 5.2.

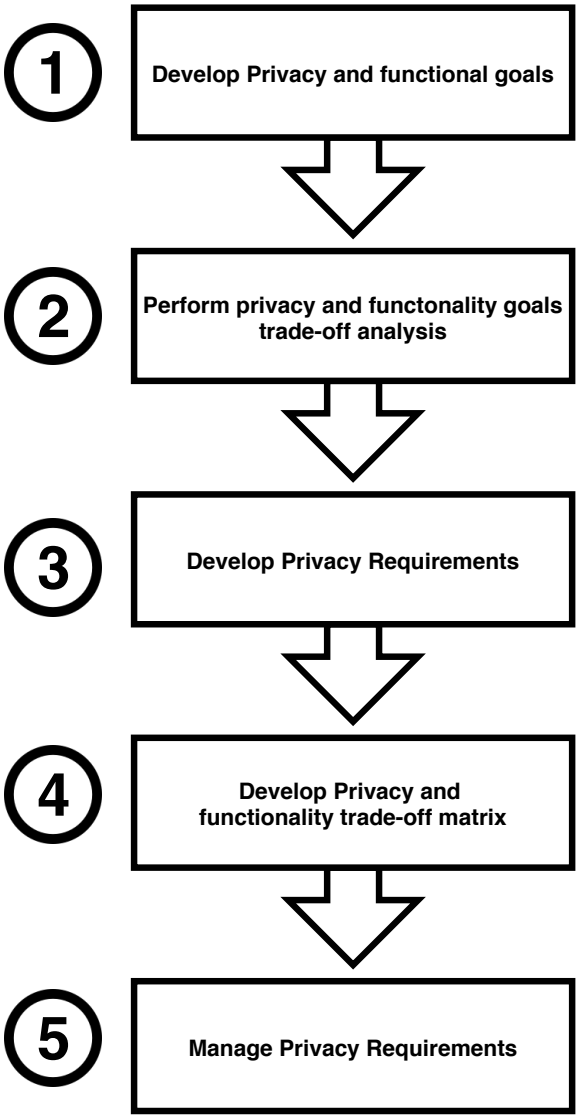


Figure 5.2: TRANK framework process flow.

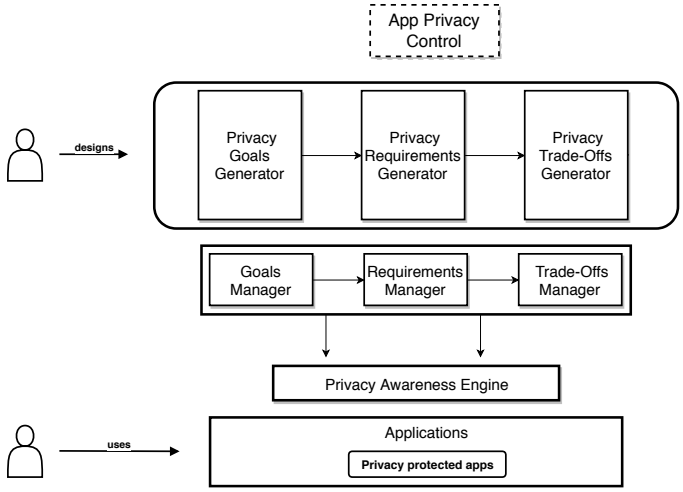


Figure 5.3: TRANK framework.

The TRANK architecture as illustrated in Figure 5.3, uses the privacy-by-architecture and privacy-by-policy design concepts to integrate privacy in mobile applications as introduced in chapter 2 of this thesis.

Step 2: Furthermore, TRANK proposes to assess and determine a functionality, performance and privacy goal trade-off analysis which aims at improving privacy modeling. The objective of this approach is to determine whether or not functionality and performance goals can be achieved even when a high level of privacy is implemented. Establishing such properties is paramount in order to determine if solutions developed enhance privacy and at the same time acceptable. The analysis identifies relationships between requirements using a requirements trade-off matrix and determines which requirements are in conflict with others. It further determines how requirements influence each other, for example by limiting access to private information in the system.

Step 3: In the third step, we propose developing privacy requirements based on privacy modeling approaches such as the LINDDUN privacy threat modeling methodology. LINDDUN has been applied by several projects e.g the BioMedBridges, RERUM, and TClouds as a privacy analysis technique. It has also been supported as a privacy risk management technique by the PRIPARE project [134]. More details of eliciting privacy requirements with the LINDDUN framework can be found in [33] [135] [136].

The LINDDUN privacy threat modeling methodology provides a systematic approach

in identifying privacy threats that may be used to compromise mobile applications. We demonstrated this approach using an empirical study to design privacy requirements in Telematics mobile applications in chapter 2. Applying LINDDUN systematically identifies privacy flaws in mobile applications and aids in fixing them before the application is developed.

Step 4: The fourth step involves reviewing and redesigning conflicting goals and requirements using a privacy requirements trade-off matrix. This step is the most challenging step as goals and requirements have to be refined in order for complex goals to produce alternative requirements. The major objective of this phase is to achieve a balance between the intrusiveness of privacy requirements to functional requirements and performance goals.

Step 5: Finally, privacy requirements are managed by a requirements management tool e.g. IBM Rational Doors [137] or Irise [138] to ensure that they are properly managed in the software development life cycle. Due to the constant changes that are involved in mobile applications, there is a high volatility of conflicts in the requirements that often need to be resolved. High volatility in software requirements demands for a quick and stable requirements management approach. Requirements management is referred to as; "an iterative set of activities that help ensure that elicitation, documentation, refinement, and changes of requirements is adequately dealt with during a lifecycle, with a view toward satisfying the overall mission or need in a quality manner and to the customers' satisfaction." [139]. Figure 5.4 illustrates how we utilized the IBM management requirements tool [137] to design and manage requirements for the telematics insurance mobile application in the course of this research. As architectures are based on the requirements that define a mobile application, changes in the requirements subsequently demand for critical architecture changes and this can result into delays or failures of mobile application software [140]. Therefore, proper requirements management is mandatory to handle requirements volatility in mobile applications.

The next section elaborates on the privacy modeling of eHealth applications and Telematics application using TRANK.

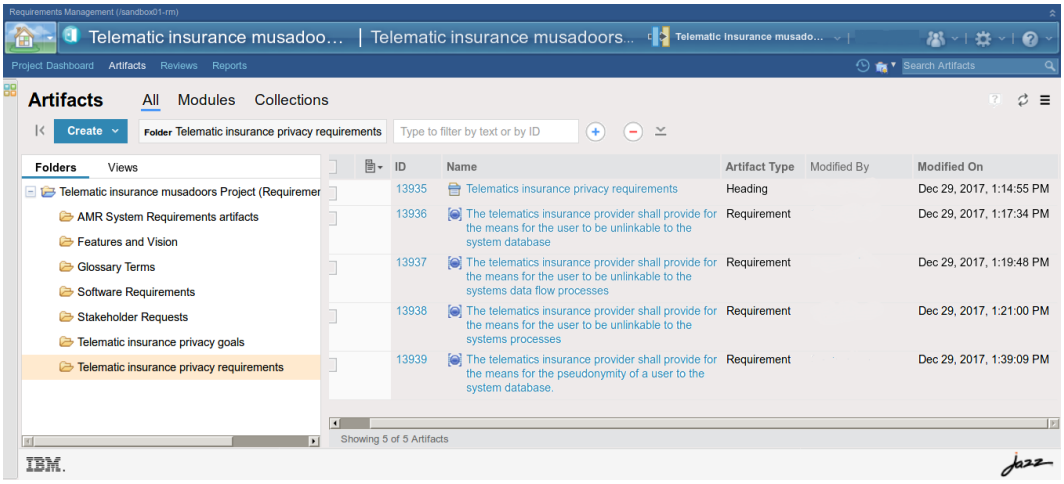


Figure 5.4: Privacy requirements management

5.2 TRANK based design for eHealth applications

5.2.1 Privacy modelling with TRANK for eHealth applications

As mentioned previously from our analysis, we observe that, in practice, modern applications are not keen on integrating multiple requirements demands simultaneously. Thus, sacrificing some of the crucial requirements like privacy requirements in turn for functional and performance requirements. This section employs TRANK and its constituent steps to model privacy using a running example of the eHealth application and a Telematics application.

In order to demonstrate the feasibility of TRANK, we modelled and generated privacy goals to address privacy intrusiveness in the eHealth application as presented in the TRANK framework. LINDDUN concepts were employed to design privacy requirements as proposed in [135]. One way to minimise privacy invasiveness is to reduce the amount of data that is collected. For example, [141] showed that the use of large volumes of data could lead to the manipulation of Personal Identifiable Information(PII). Thus, a privacy trade-off analysis was employed to reduce the amount of data collected. The privacy trade-off analysis is used to generate a design solution that best suits the applications privacy goals and requirements. It also, evaluates alternatives that can be used to produce better privacy-aware solutions.

There are various actors who play a big role in the mobile heart rate application which have to be identified. These include; the patient, the physician, system engineers, the heart beat measuring device, policy makers and shareholders among others.

These key actors play a major role in the design of a privacy aware mobile application. The interplay of the actors has to be designed in such a way that no privacy threats arise when medical data is transferred from the patient to the physician. This example shows that in order to design and maintain privacy a design methodology that integrates privacy for all actors has to be put in place. We therefore, saw the need to design a privacy framework that can be used in eHealth applications to ensure privacy is preserved and maintained at all stages of system development.

Key Actors:

- `Patient:` Uses the heart beat device to monitor his/her heart rate.
- `Physician:` Monitors the patients health and prescribes medicines as required.
- `System engineers:` rectify any problems that may occur when the telematics system is running.
- `Heart beat measuring device`: collect heart rate data for analysis by the physician.
- `System designers:` develop the heart beat monitor system from requirements specification to go live phase.
- `Policy makers:` design policies used by the heart beat monitor and other medical applications.
- `Regulators:` design rules and regulations that are used to govern eHealth devices and system applications.
- `System managers:` ensure that the heart beat monitor applications are profitable in accordance to shareholders interests.
- `Shareholders:` Generate capital for the heart beat monitor applications.

5.3 TRANK based design for Telematics applications

5.3.1 Privacy modelling with TRANK for telematics applications

Similar to the eHealth case study, first, privacy goals were designed to determine the privacy objectives required in designing telematics applications in an attempt to integrate multiple requirements demands from various stake holders and other parties involved. We further applied the LINDDUN methodology to model and integrate privacy requirements in the telematics mobile applications.

As telematics applications do collect massive amounts of both personal and location data, we performed a privacy trade-off analysis to determine which privacy requirements are required before designing the privacy aware telematics application. An example of the telematics privacy aware trade off analysis can be seen in Table 5.2.

We further identified the following key actors that are required in modelling the privacy requirements as they play a significant role because they all have different views and interests in how the application has to perform both practically and financially.

Key Actors: There are various actors who play a big role in the telematics insurance application. These include among others, shareholders who generate capital to finance application development and end users. Actors have different interests and more often their interests conflict with privacy goals and requirements. We identified the following key actors;

- `Vehicle:` used by drivers as a means of transport to desired destinations.
- `Vehicle driver:` drives the vehicle and needs insurance in case of an accident.
- `System engineers:` rectify any problems that may occur when the telematics system is running.
- `Telematics devices(blackbox)`: collect telematics data in the vehicle which is used for billing.
- `Mobile appliances:` used to connect to the system portal which monitors billing and driving style information.

Table 5.1: Modeling privacy goals.

	Privacy requirements					
Privacy goal	Unlinkability	Anonymity	Pseudonymity	Undetectability	Unobservability	Confidentiality
Instance of	satisfaction goal	satisfaction goal	satisfaction goal	satisfaction goal	satisfaction goal	satisfaction goal
Concerns	insurance company	insurance company	insurance company	insurance company	insurance company	insurance company
Informaldef	The company should maintain system unlinkability	The company should maintain system anonymity	The company should maintain system pseudonymity	The company should maintain system undetectability	The company should maintain system unobservability	The company should maintain system confidentiality

- `System designers:` develop the telematics system from requirements specification to go live phase.
- `Policy makers:` design policies used by the telematics applications.
- `Regulators:` design rules and regulations that are used to govern the telematics system applications.
- `System managers:` ensure that the telematics applications are profitable in accordance to shareholders interests.
- `Shareholders:` Generate capital for the telematics insurance applications.

5.4 Privacy goals trade-off analysis

Privacy goals are derived with the main task of the insurance company to provide maximum privacy for its customers. The privacy properties used to derive privacy goals are unlinkability, anonymity, pseudonymity, undetectability, unobservability and confidentiality. We employ The KAOS [30] formal definition of goals to model privacy goals as shown in Table 5.1.

In order to achieve the most efficient goals, a privacy goals trade-off analysis is conducted. The privacy trade-off analysis is an approach of evaluating the trade-off between privacy, system functionality and performance. The first step in the trade-off analysis is to generate functional, performance and privacy goals. We define functional goals as goals which describe the features and services that the application has to deliver. Performance goals as goals that describe the operational activity and how end users easily perceive the application. This is followed by determining the attack scenarios and weak points of the system to establish the weaknesses of the system. After establishing

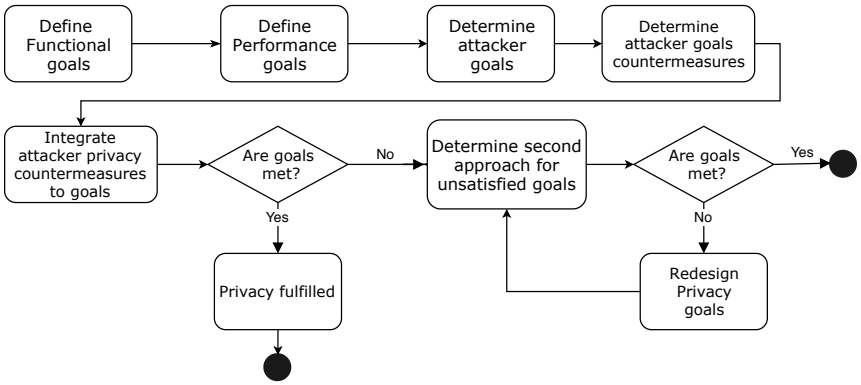


Figure 5.5: Privacy goal trade-off analysis model

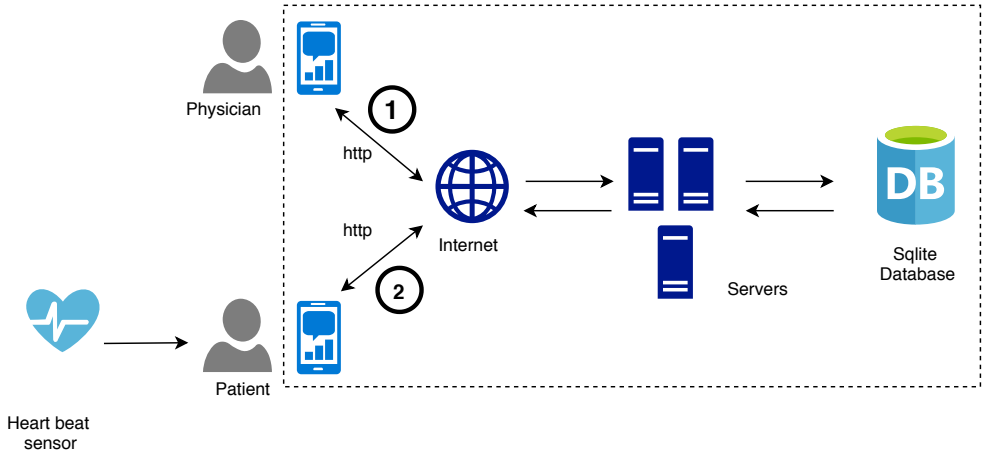


Figure 5.6: EHealth Heart rate monitor data flow diagram.

the weaknesses, solutions to these weak points are generated and countermeasures and solutions are generated. This step is followed by making a decision on finding out if the privacy and functionality goals are met. If the functional and privacy goals are fulfilled the design phase is considered complete. If they are not met, alternatives to the proposed goals are generated and a second approach to the first design is proposed.

Goals are then redesigned to meet the new design and the procedure is repeated until an acceptable solution is achieved that balances the privacy goals, functional and performance goals as illustrated in Figure 5.5.

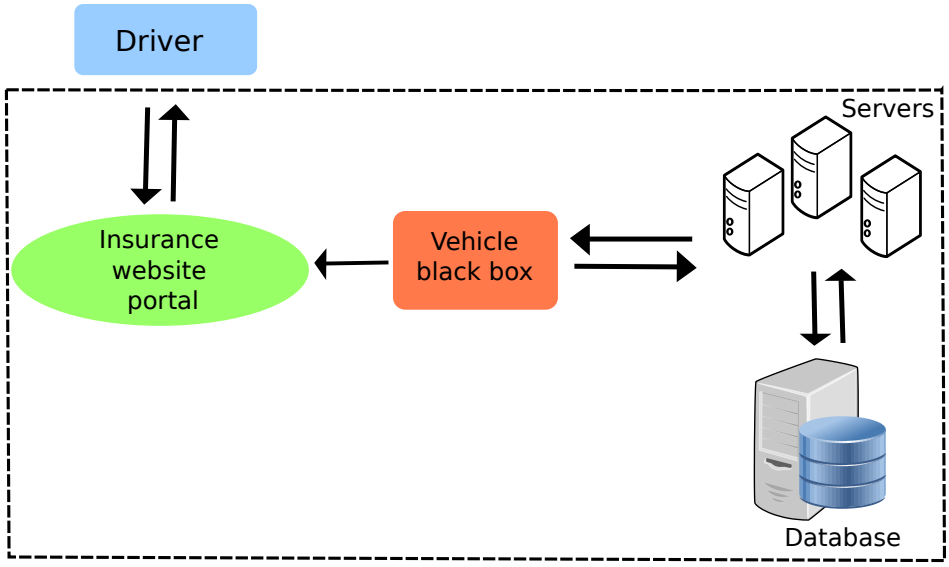


Figure 5.7: Telematics insurance data flow diagram

5.5 Privacy requirements design for mobile applications

The next step involves generating privacy requirements that are required for implementing privacy in the telematics insurance application.

Privacy requirements are derived based on the LINDDUN privacy threat model as in [135]. This task involves specifying privacy threats based on privacy properties as mentioned earlier, which include: unlinkability, anonymity, pseudonymity, undetectability, unobservability and condidentiality. The initial step of the LINDDUN methodology is to define data flow diagrams as illustrated in Figure 5.6 for the eHealth heart beat monitor application and Figure 5.7 for the Telematics application.

We use the LINDDUN methodology to determine threats using threat trees. Threat trees aid in showing attack methods and design weak points that could be used by an adversary. Figure 5.8 shows how a threat tree is designed. We refer the reader to the LINDDUN website for the current threat trees. After generating the threat trees we derive the privacy requirements and measures that can be used to mitigate the derived threats and find potential solutions to the threats [135] [136].

Privacy requirements trade-off matrix:

In this step a matrix is generated to determine the effect of privacy requirements

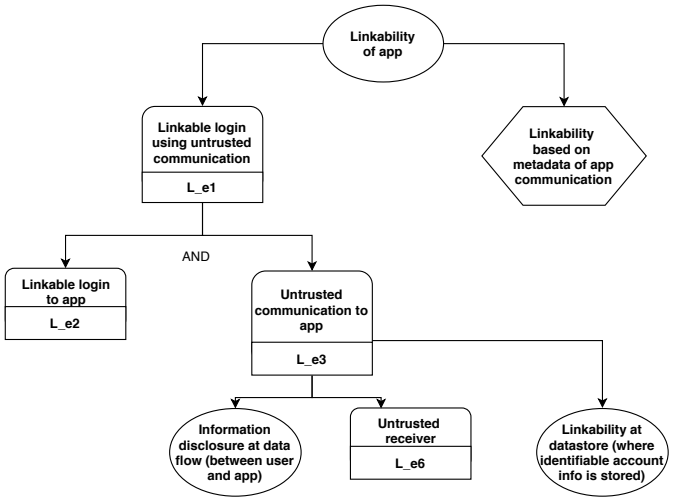


Figure 5.8: Example of a Threat tree

Table 5.2: Privacy functionality matrix template

	Unlinkability	Pseudonymity	Unobservability	Confidentiality	Policy compliance	Performance	Usability	Reliability	Scalability	Interoperability
Unlinkability		+				-	-	-	-	-
Pseudonymity					+					
Unobservability										
Confidentiality					+					
Policy compliance										
Performance	-	-	-							
Usability		-								
Reliability					+	+	+			
Scalability										
Interoperability	-	-	-							

to other functional requirements. Table 5.2 presents a template of how the matrix is generated. The various requirements may contribute negatively (-) or positively (+) to other requirement parameters. Choosing alternative privacy solutions must be done to determine how system functionality has to be maintained without compromising privacy.

We present a running example of generating the privacy functionality matrix for telematics as shown in Table 5.3 and for the eHealth application in Table 5.4. The matrix comprises of two main axes. The applications components are represented on the y-axis

Table 5.3: An example of a privacy functionality matrix for Telematics application

	Unlinkability	Pseudonymity	Unobservability	Confidentiality	Policy compliance	Performance	Usability	Reliability	Scalability	Interoperability
Drivers	5	8	7	6	2	1	4	3	9	10
Web portal	8	6	10	7	5	1	4	2	3	1
Vehicle telematics device	5	7	9	10	6	4	3	1	8	2
Back end servers	5	4	6	3	8	1	2	7	9	10
Database	6	7	10	9	5	1	4	3	2	8
Address	-	-	-	-	-	-	-	-	-	-
Email	2	3	1	4	5	6	7	8	9	10
Phone	-	-	-	-	-	-	-	-	-	-

and the privacy components are represented on the x- axis. In the telematics example, the key components are the drivers, web portal, the vehicle telematics device (also known as the blackbox), back end servers and the data base. In addition to this we include the main means of communication that are used to communicate to the users i.e email, phone and lastly the address used in the companies documents. Privacy preservation of these components plays an important role in telematics insurance and therefore, should be modeled accordingly using the privacy functionality matrix. The privacy properties are rated according to the application component. For example, the Drivers and Performance are allocated a value of 1 and are rated as the major property to take into consideration when designing the application while Interoperability is allocated a grade of 10 and considered as the least property to be considered.

In addition, an analysis of the telematics system components is done based on the privacy components to determine how they are affected in the matrix as shown in Table 5.3. Similarly, Table 5.4 presents how the privacy functionality matrix for eHealth applications are generated using the heartbeat monitor application based on the data flow diagram presented in 5.6.

Table 5.4: Privacy functionality matrix for eHealth Heart beat monitor application

	Unlinkability	Pseudonymity	Unobservability	Confidentiality	Policy compliance	Performance	Usability	Reliability	Scalability	Interoperability
Heartrate	5	6	7	4	8	1	3	2	9	10
Medical condition	4	3	5	2	6	1	7	8	9	10
Year of Diagnosis	3	2	4	1	7	5	6	8	9	10
Medication taken	5	4	6	3	8	1	2	7	9	10
Mood	-	-	-	-	-	-	-	-	-	-
Address	-	-	-	-	-	-	-	-	-	-
Email	2	3	1	4	5	6	7	8	9	10
Phone	-	-	-	-	-	-	-	-	-	-

5.5.1 Privacy requirements management for mobile applications

In this thesis we employed, IBM rational doors which is one of the key requirements engineering tools used in the industry. We chose it for our privacy modeling and management process because its well structured and highly used for the management of requirements in the industry. IBM DOORS next generation (DNG) uses *Artifacts* as the initial object when starting a project. Artifacts can be any object created in the project for example, a requirement, a heading, a diagram or an image.

5.6 TRANK Framework Design Evaluation

5.6.1 Ehealth results analysis

We initially investigated the relationship between functionality, privacy policy and data collection in eHealth apps. In the following, we present our results analysis of the examined apps based on our observation of functionality followed by results on privacy policy and an analysis on data collection. The aim of this analysis was to determine how the interplay of these features determines the design and development of current apps.

Functionality: Our analysis was based on the five categories mentioned earlier; fitness, cardiology, diabetes, weight loss, depression and physician apps. As the functionality of

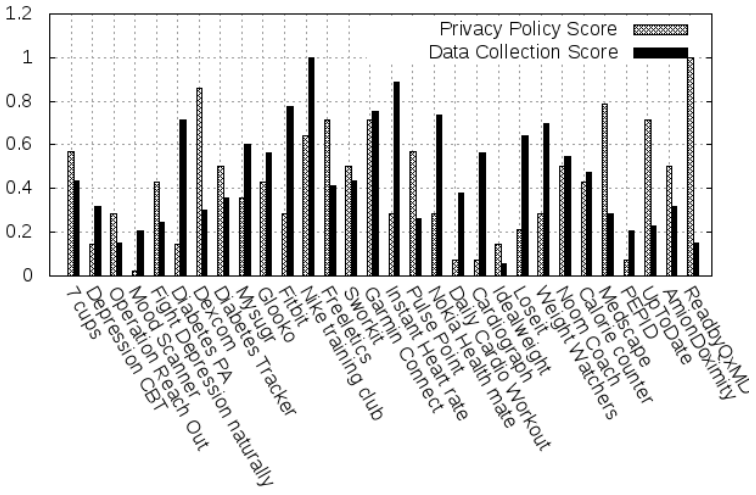
these apps varies for each app and category, we found that fitness apps had the most functionality features, followed by cardiology apps. In particular, Fitbit and Nike training had more than 50% functionality features compared to the other surveyed fitness apps. Among the noticeable features were Fitbits' ability to track down all calories a user burned. These can be used to challenge family and friends according to their sport activities. Diet routines are also recorded. Meals eaten and the number of calories taken in a day, the number of steps taken and a record of personal performance details are all collected. Fitbit also has a sleep well feature used for sleep tracking. Its used to track the different stages of sleep and also comes with a sleep schedule reminder. In contrast, diabetes and physician apps had the least functionality scores. These apps were found to be precise without a lot of functionality. For instance, Medscape an app used to aid physicians on for example, determining medications and dosages has limited additional functionality features.

Privacy policy: Most of the apps downloaded from the Google Play Store have a privacy policy on the Google Play website. This has been mainly enforced by the California Online Privacy Protection Act (CalOPPA) [142]. CalOPPA states that all applications that collect user personal information have to include a privacy policy. The privacy policy should state how applications gather user data and how its shared. One of the main objectives of this study was to find out if the privacy policies presented on the Google Play site are in sync with the privacy policies in the app and on the companies site. Our findings, however, found that only 26% of the surveyed apps had a privacy policy link presented during signup when this study was taken. For example, Depression CBT Selfhelp Guide, 7 cups anxiety, Instant heart rate, Loseit!, all had no privacy policy links in the app during sign up. In addition, 23% of all apps did not have a privacy policy on the companies website. Sites like Instant heart rate, Ideal weight had no readily accessible privacy policy on the company website. 43% of the apps surveyed reported the national laws to be followed, while 48% included the possibility of deleting personal information from their systems. 90% shared data with third parties and 87% reported that they collected data from external sources. Several apps used third party services

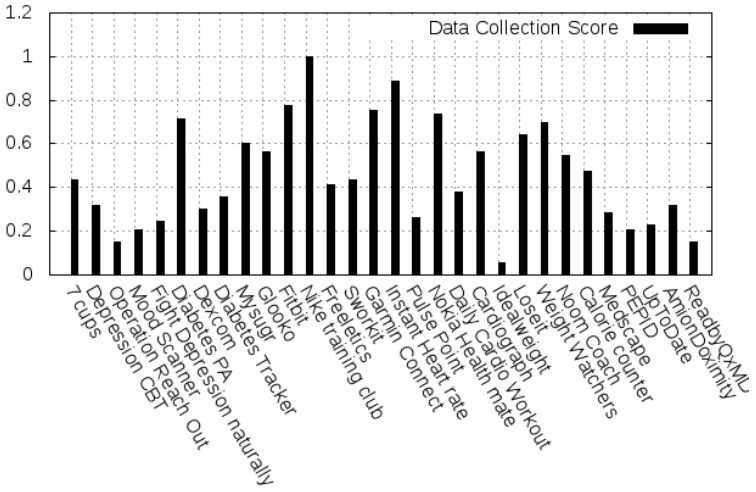
such as facebook for logging into the apps or uploaded the data to facebook. For example, Mysugr encourages users to upload their sugar levels data to facebook along with the mysugr app and their photos [143].

Data Collection: Data collection scores were investigated based on the permissions used, the data collection statements mentioned in the data privacy policies and data collected within the apps. Permissions continue to be the major data collection funnels in current mobile applications. The detailed meaning of what the permissions mean is beyond the scope of this study. However, we will highlight on the different app permissions that are maily used in the surveyed apps. Our analysis found that 94% of apps *required full network access.* 87% of the apps had access to the *read contents of the USB storage* and *modify or delete storage of your USB storage.* 80% required the *receive data from internet* and *view network connections permissions.* 47% required access to location. We found that among the least used permissions are *manage document storage* and *uninstall shortcuts* which were 3% for all surveyed apps. To expand our analysis we analysed personal data collection key words that were mentioned in the privacy policy. 63% of the surveyed apps stated that they collect the email address. 57% recorded the users name while 56% the IP address. 3% of the surveyed apps collected the Tax ID number and employer identification number. Physician apps included explicitly personal information about the data collected in their privacy policies. Some of the data mentioned are for example, medical condition (13%), mood (7%), heartrate (7%) and insurance provider (3%). In summary, Nike training club reported the most data collection score followed by Instant heart rate and Fitbit apps respectively. *Relationship between data collection and privacy policy:* Our analysis was extended to investigate the relationship between privacy policies and data collection. We hypothesize that the data collection scores are equal to the privacy policy scores. In other words, the data collection performed in the apps has to be in sync with the information stated in the privacy policies.

However, the overall findings indicated inconsistencies in how data is collected and how the privacy policies are defined. Most privacy policies do not mention which data is collected and what its used for. Some privacy policies were missing both on the company



(a) Privacy Policies vs Data collection in eHealth apps



(b)

Figure 5.9: Data collection score in eHealth apps

websites and in the app. Most of the apps had a link to the privacy policies in Google Play Store but some policies varied from those in the company websites. Also, more data was collected in the apps than what was reported in the privacy policies. The data collection scores therefore, are not in sync with the information given in the privacy policies.

We observed that fitness apps collect more data at the app level than typical medical apps such as Diabetes apps as shown in Figure 5.9 (a) and Ideal weight and Operation

reach out apps reported the least data collection scores as shown in Figure 5.9 (b).

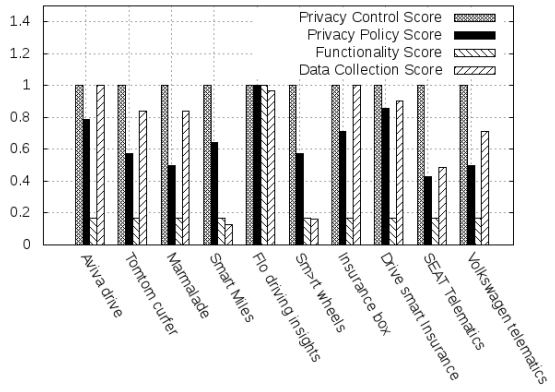
Much as the functionality of most of the apps need permissions for them to correctly work, for example a fitness app needs location information in order to determine a users mobility, users should be aware that their location information is being monitored. They should be able to have a right to opt-in and opt-out of an apps feature. Little has been done in most of the apps to allow this activity. Occasionally, there is an on and off button to choose from without explicitly explaining to the user the consequences of turning off the permissions. This is done while downloading the app which was introduced with the Android 6 version. However, more information is required on what the permissions are able to do. For example, location information permissions may lead to continuous tracking and profiling even when an app is not in use.

Therefore, there is a need to inform users about the actual data that is being collected. Our findings thus are a stepping stone in finding solutions to these gaps.

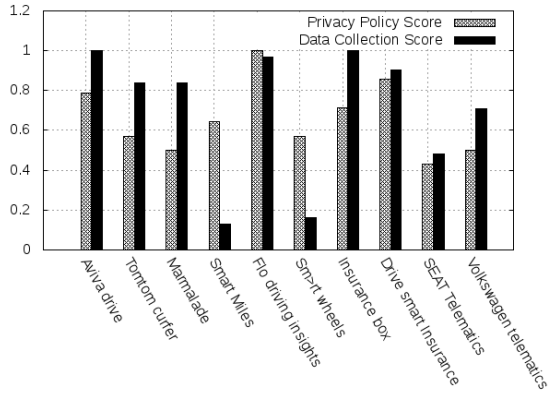
5.6.2 Telematics results analysis

Next, telematics apps were investigated, namely, the relationship between privacy policies, privacy control, functionality, and data collection as illustrated in Figure 5.10 (a). We further analysed the relationship between data collection and privacy policies as shown in Figure 5.10 (b).

The functionality score in telematics apps was almost the same in all surveyed apps. However, our analysis found that although almost all telematics insurance companies have the same functionalities, the data collection scores differ. Companies continuously collect data even when its not required for better functionality and performance. We observed that Insurance box and Aviva Drive had the highest data collection scores. Furthermore, the privacy policy scores varied from telematics app to app. Of the surveyed apps only 20% had a privacy policy link presented during signup although all apps collected data from external sources. For example, Aviva drive, Marmalade, Sm>rt wheels, Insurance box did not have a privacy policy link during signup. 40% specified the national laws which were used in their privacy policies. In addition, only 30% reported on the ability



(a) PCS, PPS, DCS, FS in V2X telematics apps



(b) PCS vs DCS for V2X telematics apps

Figure 5.10: Relationship between PCS, PPS, DCS and FS in V2X telematics apps

of users to delete their personal information. These observations indicate that privacy policies in V2X telematic apps still need improvement compared to ehealth apps. Much as the privacy policies did exist, especially on the Google Play Store website, most of the companies did not include them in the company websites.

Telematics applications highly rely on data collected in the vehicle to compute insurance policies. A number of sensors in the vehicle e.g. GPS sensors, G-force sensors are used to measure a drivers speed, brakes and acceleration. Insurance companies determine how well a user drives based on such sensor data. Therefore, V2X telematics insurance involves a lot of data collection. However, users are not explicitly told of how this data is collected and how its used and processed. Based on the GPS sensor monitoring functionality, it

follows that the most used permission in V2X telematics apps is *approximate and precise location*. Our analysis found that only 20% of the surveyed telematic insurance companies included this information in their privacy policies.

In general, there has been a steady improvement in apps on how permissions are accessed. However, from our analysis more emphasis has to be done on informing users what exactly the app permission means and the functionality it performs.

5.6.3 Evaluation of the TRANK framework

To evaluate the TRANK framework we employed it to two use cases the eHealth and V2X telematic applications. Our assessment characterizes the improvement of privacy analysis results produced by TRANK as well as the completeness associated with executing the framework. We focused on how TRANK affects data collection and functionality.

5.6.4 Impact of TRANK on Data collection

To properly assess the accuracy of TRANK on the data collection score we analysed the values of data collection scores before and after using the TRANK framework as presented in Table 5.5 and 5.6.

We hypothesized that the data collection scores before using TRANK are less than or equal to the data collection scores after using TRANK.

H0: Data collection score before using TRANK are less than or equal to the data collection scores after using TRANK.
H1: Data collection score before using TRANK are greater than the data collection scores after using TRANK.

$$\begin{aligned} H_0 : \mu DCS_b - \mu DCS_a \leq 0 \\ H_1 : \mu DCS_b - \mu DCS_a > 0 \end{aligned} \tag{5.1}$$

We ran a paired samples t-test with 95% confidence interval on eHealth applications for data collection scores after employing TRANK (Number of apps = 30, Mean = 15.00,

Table 5.5: eHealth app scores before and after using TRANK

Mobile App	Privacy Control Score	Privacy Policy Score	Data collection store 1	Data Collection Score 2
Depression				
7 cups	0.958504297	0.571428571	0.433962264	0.086956522
Depression CBT	0.470588235	0.142857143	0.320754717	0.086956522
Operation Reach Out	0.276470588	0.285714286	0.150943396	0.163913543
Mood Scanner	0.12846572	0.024354876	0.20754717	0.060869566
Fight Depression naturally	0.176470588	0.428571429	0.245283019	0.073913043
Diabetes				
Diabetes PA	0.411764706	0.142857143	0.716981132	0.265217391
Dexcom	0.235294118	0.857142857	0.301886792	0.104347826
Diabetes Tracker	0.216359793	0.5	0.358490566	0.104347826
Mysugr	0.529411765	0.357142857	0.603773585	0.082608696
Glooko	0.024957305	0.428571429	0.566037736	0.395652174
Fitness				
Fitbit	0.882352941	0.285714286	0.773584906	0.434782609
Nike training club	0.941176471	0.642857143	0.992495633	0.230434783
Freeletics	0.588235294	0.714285714	0.415069434	0.178260878
Sworkit	0.176470588	0.594857394	0.433962264	0.291304348
Garmin Connect	0.176470588	0.714285714	0.754716981	0.218073913
Heart				
Instant Heart rate	0.647058824	0.285714286	0.886792453	0.143478261
Pulse Point	0.411764706	0.571428571	0.264150943	0.160869565
Nokia Health mate	0.294117647	0.285714286	0.735849057	0.996554329
Daily Cardio Workout	0.117647059	0.071428571	0.377358491	0.082608696
Cardiograph	0.529411765	0.071428571	0.566037736	0.265217391
Weightloss				
Idealweight	0.117647059	0.142857143	0.056603774	0.002382382
Loseit	0.529411765	0.214285714	0.641509434	0.113043478
Weight Watchers	0.352941176	0.285714286	0.698113208	0.034782609
Noom coach	0.235294118	0.5	0.547169811	0.178267087
Calorie counter	0.117647059	0.428571429	0.471698113	0.278286087
Clinician				
Medscape	0.128473649	0.785714286	0.283018868	0.147826087
PEPID	0.112937465	0.071428571	0.20754717	0.147826087
UpToDate	0.328173947	0.714285714	0.226415094	0.191304348
AmionDoximity	0.201928374	0.5	0.320754717	0.047826087
ReadbyQxMD	0.380210384	0.993837346	0.150943396	0.160869565

Table 5.6: Telematics app scores before and after using TRANK

App	Privacy control score	Privacy policy score	Functionality score	Data Collection score	Data Collection Score2
Aviva drive	0.763987654569	0.7857142857142857	0.16666666666666667	1	0.6837463234445 31
Tomtom curfer	0.730234567234	0.5714285714285714	0.16666666666666667	0.8387096774193548	0.4322565692374 83
Marmalade	0.983434058332	0.5	0.16666666666666667	0.8387096774193548	0.3849397204756 34
Smart Miles	0.890937553645	0.6428571428571429	0.16666666666666667	0.1290322580645161	0.0232425128947 52
Flo driving insights	0.960835374372	1	1	0.9677419354838709	0.5935632048563 5
sm>rt wheels	0.572345983758	0.5714285714285714	0.16666666666666667	0.1612903225806451	0.0829384572027 63
Insurance box	0.825635472547	0.7142857142857143	0.16666666666666667	1	0.5923756297463 92
Drive smart Insurance	0.934873642904	0.8571428571428571	0.16666666666666667	0.9032258064516129	0.4823942748375 63
SEAT Telematics	0.783125364826	0.4285714285714286	0.16666666666666667	0.4838709677419355	0.2948572367236 58
Volkswagen telematics	0.847127947254	0.5	0.16666666666666667	0.7096774193548387	0.3802457633592 79

Table 5.7: Paired Samples Statistics for eHealth after using The TRANK framework.

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	DCS Before	24.23	30	12.75	2.327
	DCS After	9.233	30	4.89	0.89

Table 5.8: Paired differences for eHealth after using the TRANK framework

Paired Differences									
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Data collection score before Data collection score after	15.00	10.33	1.89	11.14	18.86	7.59	29	0.00

Standard Deviation = 10.32) conditions t(29) = 7.95 p-value = 0.00 as shown in table 5.7 and Table 5.8. We therefore reject the null hypothesis and accept the alternative hypothesis.

By employing a privacy goals trade off analysis, only the goals that are required for the application to perform effectively were used to determine the data collection score. We came to the conclusion that the scores were reduced thus improving the privacy score. This was clearly depicted especially in the case of fitness apps which collect a lot of data that is not required for an effective app performance. A good example of this phenomenon was although Fitbit and Nike training apps have almost similar functional goals; that is to provide a healthier active life to their users, Nike training collected more data.

Using the TRANK framework to design such apps reduces the amount of data collected thus improving privacy preservation in the application.

Furthermore, we observed that using TRANK significantly reduces the amount of data collection. As a result only the most required goals are used to determine which permissions are to be accepted.

On the whole, the purpose of the app was not correlated to the amount of data collection performed by the application. That is, although some apps had the same functionality, their data collection scores varied.

5.6.5 Impact of TRANK on Functionality

We found significant differences when utilizing TRANK on both telematics apps and eHealth apps as the inbuilt functionality of each app differs. For example, although telematics apps have almost similar functionalities the impact of using the TRANK framework differs according to each app based on the companies business model. This is mainly because some companies use vehicle diagnostics hardware which uses the On-Board diagnostics II port, while others use devices such as smart phones or Bluetooth solutions to collect and analyze telematics data. Thus, the data collection differs with each app and therefore using the TRANK framework results into different results. We found that employing TRANK on vehicle diagnostics hardware telematic solutions and bluetooth solutions resulted into lower functionality scores compared to smartphone solutions.

Using TRANK significantly improves privacy in both eHealth and V2X telematics apps. TRANK employs a privacy goal trade-off analysis that checks for privacy implementation at the initial stages of system development. Existing privacy modelling approaches for example, the OASIS Privacy management Reference Model and Methodology [144] do not take into account the trade-off between functionality and privacy. TRANK, however, examines the requirements each application requires for both functional, performance as well as privacy preservation which is a crucial step in privacy enhancement in not only eHealth and telematics applications but in all future mobile applications to be developed.

5.7 Summary

In this chapter, we propose the TRANK framework to design privacy aware mobile applications. This work is motivated by the existing privacy breaches that evolve as a result of massive data collection in current mobile applications. In order to provide a more robust privacy aware design of mobile applications, we proposed the generation of privacy goals at the initial stages of development followed by the elicitation of privacy requirements. To ensure that data collection is minimized a privacy goal matrix and a privacy trade-off analysis is proposed. This enhances privacy preservation by reducing the

amount of data collected and ensuring that only data required for system functionality is stored. This section further integrates the TRANK framework based design in two case studies, the eHealth and Telematics applications case studies. First, we explore the implementation of TRANK in the eHealth domain using a data set comprising of 30 eHealth mobile applications. Data collection and extraction in these apps was investigated to determine existing privacy challenges. Next, telematics insurance mobile applications were analyzed and data collection and extraction were investigated to determine the state-of-art of privacy enhancement in these domains. A privacy matrix was introduced to tackle the problem of trade-offs between functionality and privacy.

Chapter 6

TRANK based application development

This chapter demonstrates the applicability of TRANK in order to develop mobile applications using the eHealth and Telematics insurance applications. We focus on these two domains because of their evasive nature and because they constantly collect sensitive data which requires a high level of privacy protection.

Our framework aims at assisting mobile application developers in designing privacy-aware applications while considering the trade-offs that occur between system functionality and performance. Our framework implementation TRANK, therefore, provides the following:

1. Aids in generating a privacy trade-off analysis for the application under development.
2. Integrates user privacy settings in accordance with the application used and privacy features chosen.
3. Aids in defining user-specific privacy features according to the activities taken.

In the following sections, we describe the general overview of the TRANK framework, its architectural components and the relationships between the different privacy components.

6.1 Introduction

Mobile applications are used for a range of online services across private and commercial domains. These domains need to be secured and resilient to face challenges of privacy leakages, privacy misconfigurations, cyber security attacks and system failures. Current mobile applications have, however, faced a lot of criticism about unauthorized and unintentional transfer of sensitive data due to misconfigured back ends, data sharing and transfer to third-party service providers. In particular, the range of beneficial opportunities and functional properties offered by mobile applications such as communication through OSNs, easy accessibility, better functionality, business enhancement, introduce a number of vulnerabilities. An indirect drawback lies within mobile applications dependency on the Internet where privacy protection has been extensively studied but still faces setbacks. The approach taken in this work relies on the TRANK framework to design privacy aware mobile applications. TRANK is a privacy aware trade-off analysis framework introduced in the previous chapter that is used to integrate privacy in mobile applications. The underlying assumption is that future mobile applications will be subjected to novel cyber security attacks, privacy breaches and anomalies for which existing solutions will be insufficient and ineffective. Therefore, the goal of this work is to develop privacy protection techniques that are particularly targeted for mobile application development and design.

The infrastructure is based on the native mobile application design as illustrated in Figure 6.1 [145].

We consider the elements that make up a mobile application, which are; the Presentation layer, the Business layer and the Data layer. The Presentation layer contains the User Interface(UI) components and the User Interface(UI) process components. The UI process components are a combination of views ad controllers used to perform the Presentation logic. The Presentation logic contains routines and instructions responsible for executing the Graphical User Interface. The Business layer is composed of the Application Facade, Business Work flows, Business Components and Business Entities. The Application Facade contains a combination of business operations and provides a single-message based

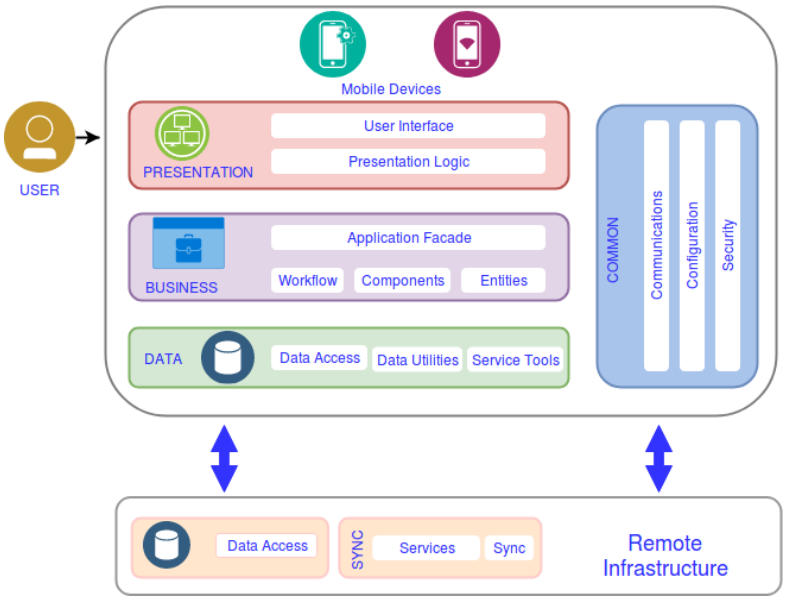


Figure 6.1: Mobile Application Architecture.

operation which can be accessed through different communication technologies. Business components on the other hand provide users with business services based on business rules. Business entities are used to transfer data between different components e.g. data from storage databases, database views, database tables. Business work flows are used to coordinate business processes in multiple steps. The Data access layer is composed of the Data Access component, the data utilities component and data service tools. This layer is used to access logic components, manipulate and transform data using helper methods and utilities. It also contains service agents that manage communication to services provided by the business component. The multi-layered architecture connects to external data sources or a remote infrastructure using synchronization connected to local device databases. Multiple database types e.g. Oracle, Microsoft SQLServer and DB2 are synced with applications using Sync services. In addition to these components, common or cross-cutting functionality such as security, communications and configuration management is used for logging and instrumentation.

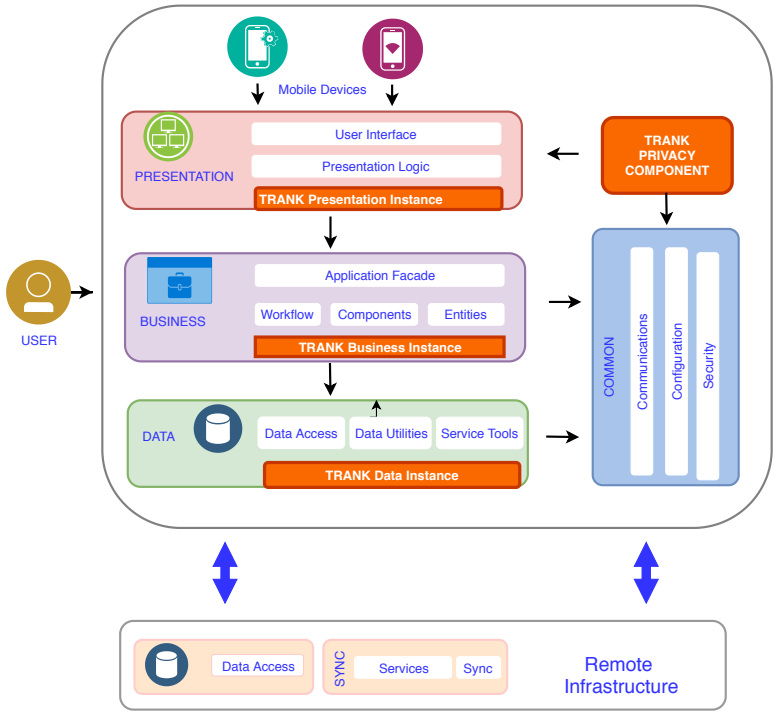


Figure 6.2: TRANK Mobile Application Architecture Design.

6.2 Gaps in developing privacy preserving mobile applications

The native mobile application architecture would ideally be free from cyber attacks, vulnerabilities and data breaches. However, we see an increase of privacy and security threats in current mobile applications. There remains a need for an efficient method that can integrate privacy in mobile applications at the initial stages of application development. Few researchers have addressed the problem of privacy challenges in mobile applications, in particular, the gradual increase of privacy-invading mobile apps.

Although the existing mobile architecture entails a common security component, unfortunately its functionality does not always guarantee effective security and privacy protection. An alternative approach to ensure privacy protection is necessary. There exists a gap in the design of the existing architecture because of the absence of a privacy protection component as seen in Figure 6.1.

This thesis closes this gap by proposing the integration of a privacy protection

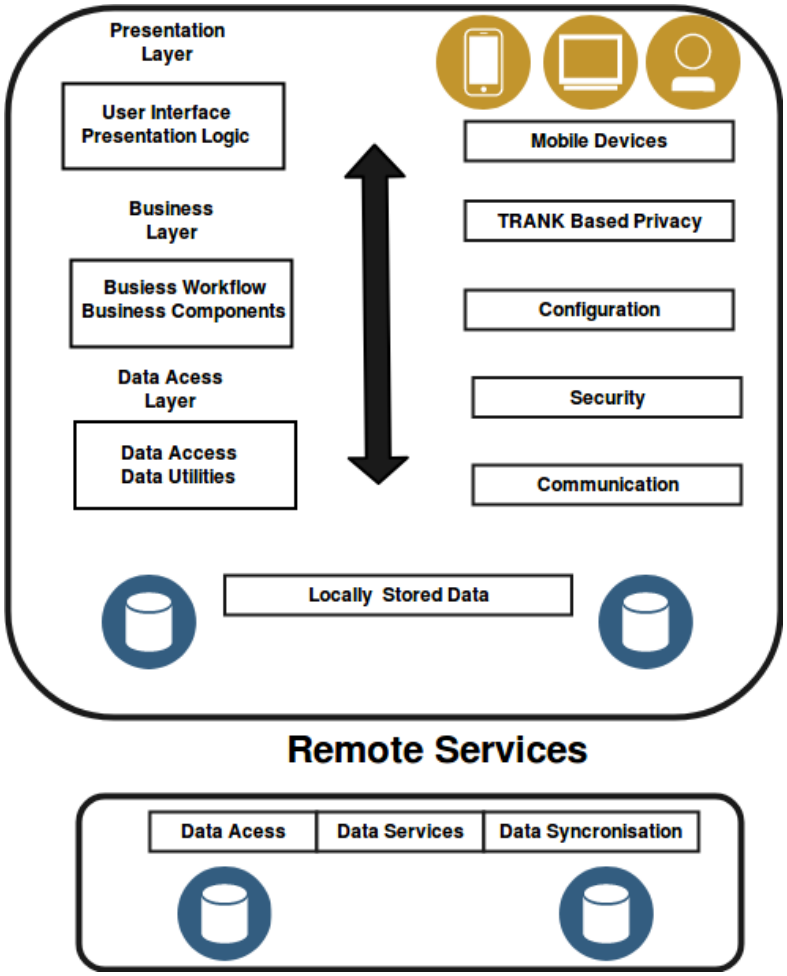


Figure 6.3: TRANK Mobile Application Architecture.

component called the TRANK Privacy Component as shown in Figure 6.2. Comparing Figures 6.1 and 6.2 shows that a privacy preservation component is crucial in an attempt to safeguard privacy in mobile applications.

In order to increase preservation of the mobile application infrastructure, we have defined a resilience architecture as illustrated in Figure 6.3. This work discusses privacy preservation using a novelty privacy detection and preservation approach that employs a privacy monitoring framework. The architecture extends the native mobile architecture by integrating a privacy aware trade-off analysis framework named TRANK to ensure privacy protection. TRANK introduces three major components within this architecture that deal with privacy preservation at the Presentation and Business layer. The components in TRANK present a foundation in which different privacy protection techniques allow the

identification of privacy threats and enable data protection. More specifically, we evaluate our approach using a privacy enhancing methodology as emulated within a prototype implementation. The major privacy preserving methodologies used are a privacy threat modeling technique, a privacy and functionality trade-off analysis technique and a data minimization technique. We have selected these particular privacy preservation and enhancing techniques as they have been identified to effectively enhance privacy preservation and reduce data collection based on the studies we performed. Our contributions in designing the TRANK prototype are as follows:

1. We introduce a privacy preservation architecture that is evidenced through an implementation of a privacy aware trade-off analysis framework across a multi-layered architecture.

2. To the best of our knowledge, our work is the first to explicitly address the aspect of trade-offs in privacy design by introducing a privacy and functionality trade-off matrix that plays a major role in designing mobile applications.

3. We provide a prototype implementation which employs "Privacy Modes" that enable end-users to have control over private data transferred to mobile applications.

6.3 TRANK based Mobile application architecture

The architecture introduced in this section is based on the native mobile application architecture comprising of a Presentation layer, a Business layer, a Data Access layer, a remote infrastructure data access layer and a common or cross-cutting component. In contrast to the native mobile application layer, we introduce a TRANK based privacy aware sub-component in the common or cross-cutting component. We emphasize on the integration of privacy in the future mobile application architecture. As indicated in the native mobile architecture a component that handles privacy development is evidently missing. Cross-cutting functionality such as configuration management, communication across multiple boundary layers and security features e.g. those against cross-site scripting are included in the architecture but no privacy functionality is inclusive. The main issue

raised in this work is the integration of a privacy preservation component that addresses the aspect of early privacy abuse prevention, privacy abuse detection and minimal PII data collection. Therefore, apart from providing a privacy trade-off analysis framework, our work further aims at enriching the native mobile application architecture with a privacy management sub-component. Hence in this work we are inspired by the works of [27], where privacy-by-architecture is emphasized. We employ privacy-by-architecture explicitly for the prevention of privacy attacks in future mobile applications, in particular those resulting from architectural loopholes and excessive data collection. One further reason to integrate privacy in mobile application architectures is to design mobile applications where privacy is by default but not responsive as in the case of current mobile applications. As a result it is possible to prevent future privacy attacks where users do not have to cater for any privacy preserving measures but the application handles them in advance without end-user interaction.

TRANK Based Privacy Component: The TRANK Based Privacy Component (TBPC) is a multi-layered component that comprises of 5 major sub-components as illustrated in Figure 6.4 namely; the Privacy Threat component, the Privacy Trade-off component, the Privacy Mode component, the Privacy Permissions Integration component and the Privacy Policies Management component. It also embodies two cross-cutting sub-components namely; the Goals component and the Privacy requirements management component.

In particular, TBPC is comprises of inner privacy control sub-components responsible for *Defending* the mobile application against privacy threats, *Determining* privacy trade-offs, *Minimizing* privacy breaches using a Privacy Mode component, *Integrating* Privacy permissions and *Management* of Privacy policies in mobile applications. And outer cross-cutting sub-components that refine the whole system by constant *Diagnosis* of weaknesses in privacy goal management and privacy requirements management. Whilst the inner privacy control component aims at privacy preservation of the mobile application, the outer sub-component aims at ensuring that the privacy goals that are generated at the initial stages of mobile application development are retained and the privacy

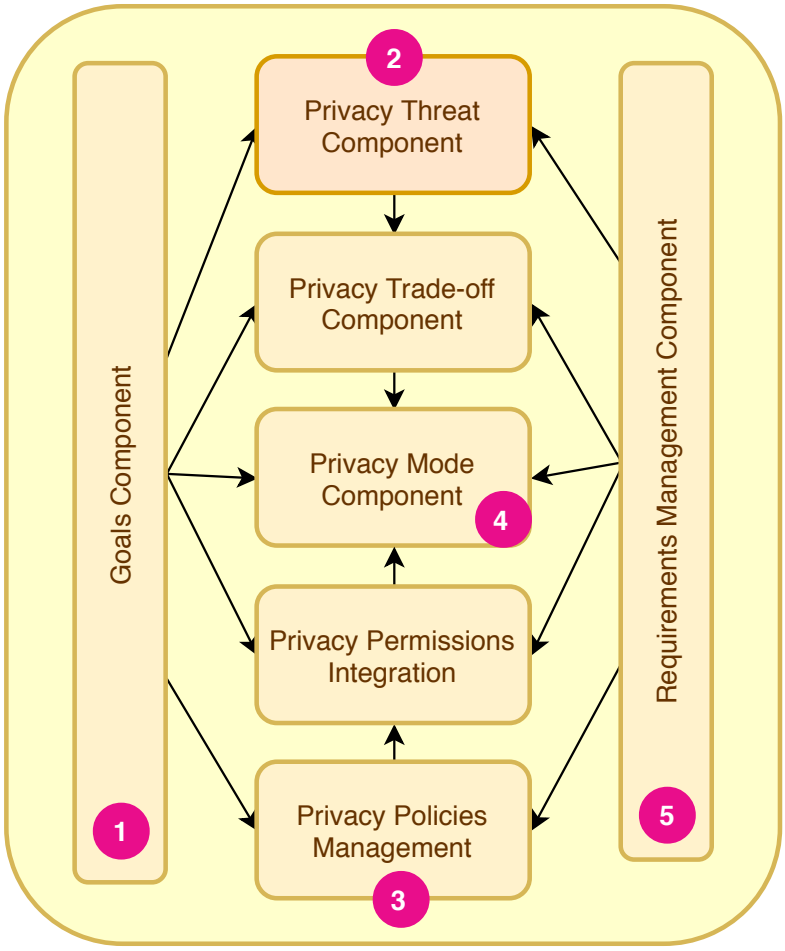


Figure 6.4: TRANK Based Privacy Component.

requirements are managed efficiently. Based on the features gathered from each individual sub-component, the TRANK Based mobile application implementation architecture is designed to enforce techniques that are capable of building privacy enhancing solutions in future mobile applications. In our implementation, privacy features are introduced in the mobile application architecture and are combined to form a privacy feature composite. Under normal operation the privacy features are run simultaneously to ensure that the TRANK based privacy module is run smoothly and efficiently. The following sections are dedicated to describing the design and steps involved in the prototype implementation in order to integrate the TRANK based mobile architecture in a running mobile application.

Data Flow:

The TRANK based mobile application used in this work is based on a native mobile

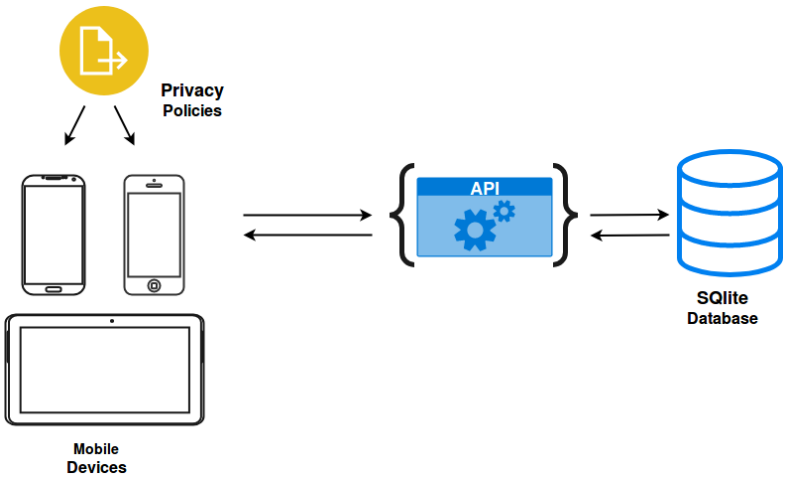


Figure 6.5: Trank Data Flow.

application which consists of a Presentation Layer with User Interface components, a Business Layer with business entities and business components, and a Data Layer comprising of Data Access Components and data utilities consisting of various devices. These range from smart phones to other portable device types like tablets running on different hardware systems and software requirements as illustrated in Figure 6.5. The resolution of the screens to be used, CPU characteristics, Memory usage, the size of the screen and its resolution have to be considered when designing mobile applications. In addition, in order to improve the UI design we emphasized on a UI where users can interact intuitively and easily. Navigating through the application both at the front-end and the back-end is crucial especially between the Presentation layer and the Data Access layer. In our implementation, the introduction of the TBPC which contains the privacy features has to be executed in a manner that end-users can easily navigate through the privacy modes. In this case the data from the TBPC e.g the Privacy policies generated from the Privacy Policy Management component are transfered to the UI from the Data Access Layer asynchronously.

The eHealth API can be found on github under the following link:

`https://github.com/schmusa/eHealth`

eHealth User Integration: Figure 6.6 illustrates an example of the data flow in an eHealth mobile application. Consider a sensor detecting a patients heart rate, using

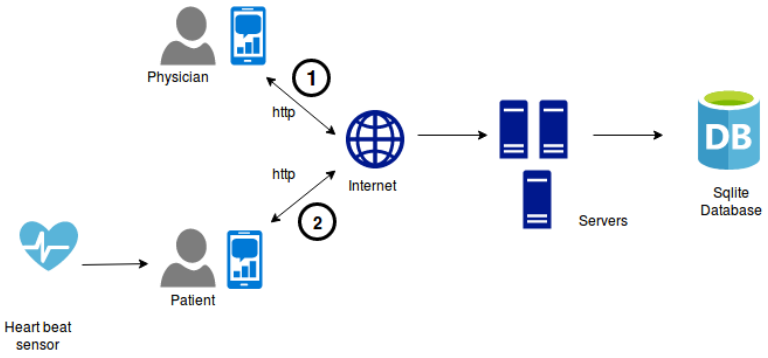


Figure 6.6: EHealth User integration.

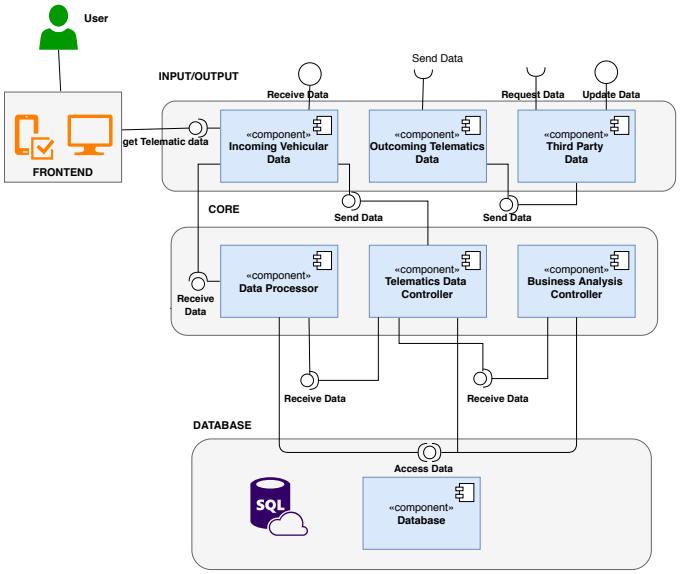


Figure 6.7: Telematics components

for example a wearable gadget or a sensor integrated in the mobile device. The patients data is transferred to the the Internet using Https protocol to the back-end servers and stored into a SQLite database. If a Physician controls a patients heart rate readings, the request is sent via the Internet to the back-end servers and a result is generated back to the Physician for proper diagnosis. Patients medical data requires a high level of security and privacy. In our implementation we introduce the TBPC which ensures that privacy preservation of the data transfered from the patient to the Physician is maintained.

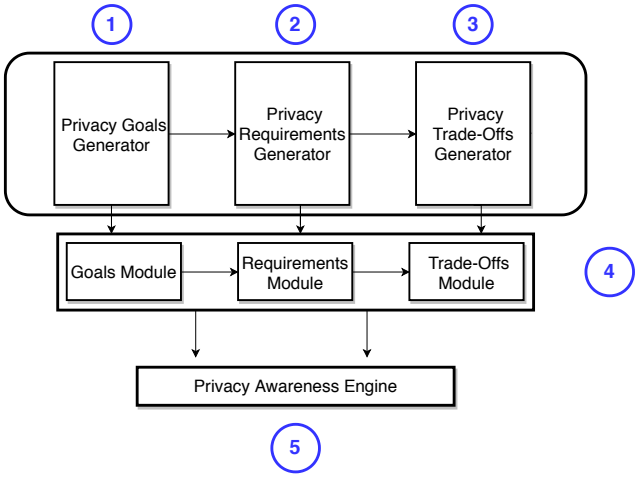


Figure 6.8: Prototype Requirements modelling using TRANK

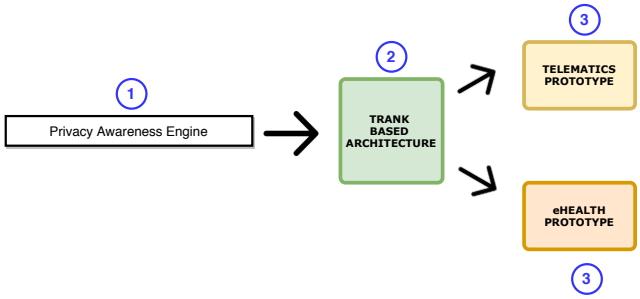


Figure 6.9: Integration of TRANK in the Prototype Architecture and Implementation

6.4 TRANK Based Mobile application demo

This section describes a mobile demo-application which focuses on improved privacy integration compared to currently available applications. The prototype was built based on the TRANK conceptual framework process. Initially, privacy requirements were generated taking into consideration the privacy goals and privacy and functionality trade-offs as shown in Figure 6.8. As a result a to privacy awareness engine is developed. The privacy awareness engine is then integrated in the TRANK based architecture. The TRANK based architecure is then used to develop the eHealth and Telematics prototype as illustrated in Figure 6.9.

We focus on the Privacy Mode Component which plays a key factor in implementing privacy in mobile applications. Its key features are so-called "Privacy Modes" which give the user control over how their personal data is used by the respective mobile application.

TRANK based Privacy modes: The TRANK framework application implementation is based on Privacy Modes. The Privacy Modes within the Privacy Mode Component (PMC) are; Maximum Data Mode, Minimum Data Mode, Users Specific Mode and the Data Obfuscation Mode. The Privacy Mode Component (PMC) in each mobile application performs privacy misuse detection based on the features gathered from the other components in the TRANK Based Privacy Component (TBPC). The PMC is in charge of coordinating and disseminating information between other components within the TBPC based on the output of the privacy abuse prevention and remediation functionality they execute. These are then used to pinpoint privacy flaws and execute privacy preserving mechanisms.

Current mobile applications allow users to choose which permissions they prefer to use. The permissions enable apps to access phones features e.g. location, storage, media e.t.c. Much as users can decide on which permissions to choose, the possibility of an application not functioning as desired is very high. For example, if a user decides to deselect location permissions, location based applications will not perform as expected. We therefore, propose privacy modes to enable users choose permissions without losing functionality. The PMC comprises of the following modes;

Maximum Data Mode: The Maximum Data Mode (MDM) is in charge of the overall data control provided by the user. It enables users to submit all the data asked for as in current applications. Herein, the application can access all the data including optional data which is not mandatory for the application to perform as expected.

Minimum Data Mode: The Minimum Data Mode (MDM) allows the user to have access to preferences mandatory for the application to perform. This means that the user can select the basic preferences without selecting other permissions that are innesential for the application .

User Specific Mode: The User Specific Mode (USM)allows users to choose which permissions preferences to have access to. Here users can grant apps permissions to access data, hardware functionality and features depending on what they prefer.

Obfuscation Mode: The obfuscation mode can also be referred to as the Data

Anonymizer Mode. This enables users to select preferences that are obfuscated or anonymised. This way if a user, for example, wants to use a location based app but does not want to show his or her exact location, anonymised location data can be selected to protect user identity and privacy.The framework is designed as a multi-tier architecture based on Privacy Modes and the above named components. The upper layers of the framework interact with the lower layers using fine grained function calls.

The application is based on a standard Android application written in the Java programming language. For easier implementation, some additional libraries have been used. Namely, several Google support libraries as well as an open source graph plotting library [1] developed by GitHub user jjoe64. Android apps are comprised of four major components; An *Activity* which is the User Interface visible to end-users, a *Service* which generates executable code in the back-end servers, a *Broadcast Receiver* which receives and broadcasts application events and a *Content Provider* which manages access to data resources in structured databases e.g. SQlite or unstructured data like flat files.

In our implementation, the mobile application collects eHealth data in this case, fitness data in form of "Sessions" and stores them locally on the user's device. It is viewed and managed via the app. Sessions which currently include the user's pulse and meta data such as timestamps and location whereby the latter is used to display workouts in a Google map.

6.4.1 TRANK based main component

The overview activity is the main component and entry point of the application. It provides the user with the most relevant information at a glance as shown in Figure 6.10. From here, it is also possible to quickly navigate to the privacy settings. The first available tab gives the user an overview of his current privacy settings e.g. his last recorded fitness sessions or data. From the recorded sessions the user can easily navigate back to the privacy settings. We integrate a *Timeline* as shown in figure 6.11 which lists all sessions stored on the device, these are further accessed via a floating action button.

[1] http://www.android-graph view.org/

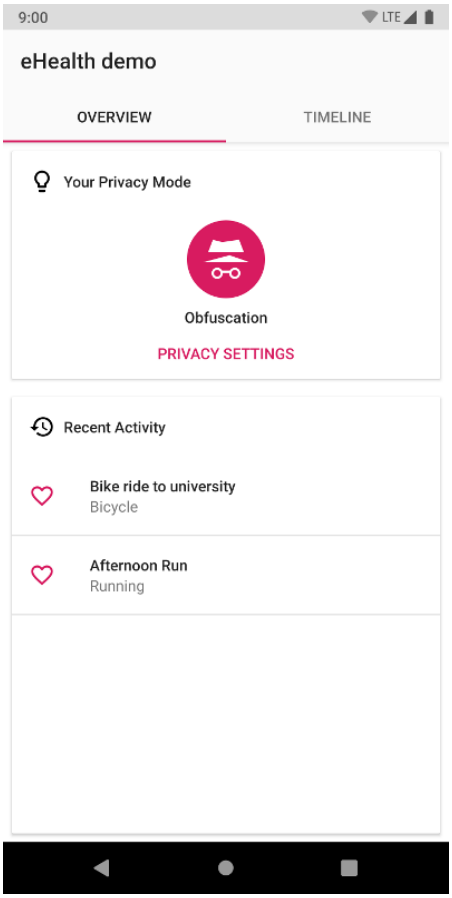


Figure 6.10: eHealth Overview.

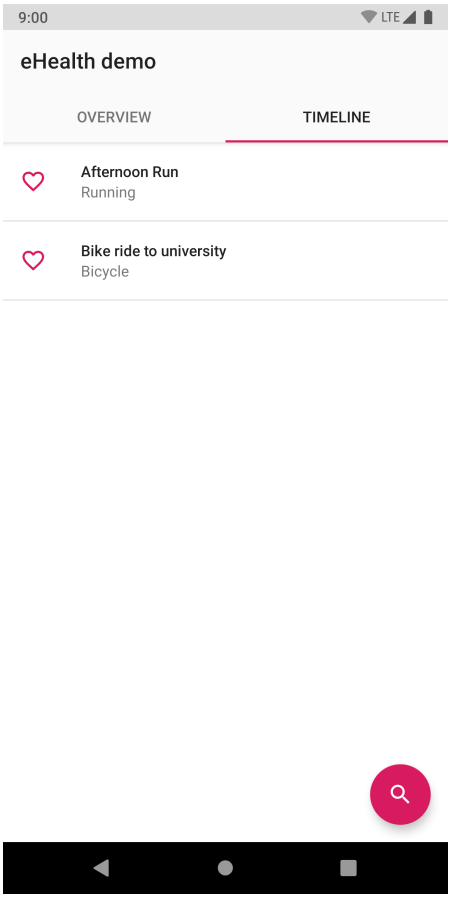


Figure 6.11: eHealth Timeline.

6.4.2 TRANK based Privacy Settings

The Data Access Layer is accessed through a *Data Activity* used to present information about a fitness session in detail as shown in Figure 6.12. It consists of an action bar at the top containing the session title, a delete button, a map, and multiple other categories for each recorded property. For simplicity and demonstration purposes, this is implemented as a user's heart rate.

Privacy settings design plays a crucial role in this work, this is where the Privacy Mode Component which is the core functionality of the application prototype is implemented. The privacy settings activity is the main point to select a privacy mode and customize its respective settings. Figure 6.13 shows a screenshot of the privacy settings. The currently available modes are "*Maximum Data*", "*User Defined*", "*Minimum Data*" and "*Obfuscation*" or "*Data Anonymizer*". The user has the option to edit additional settings

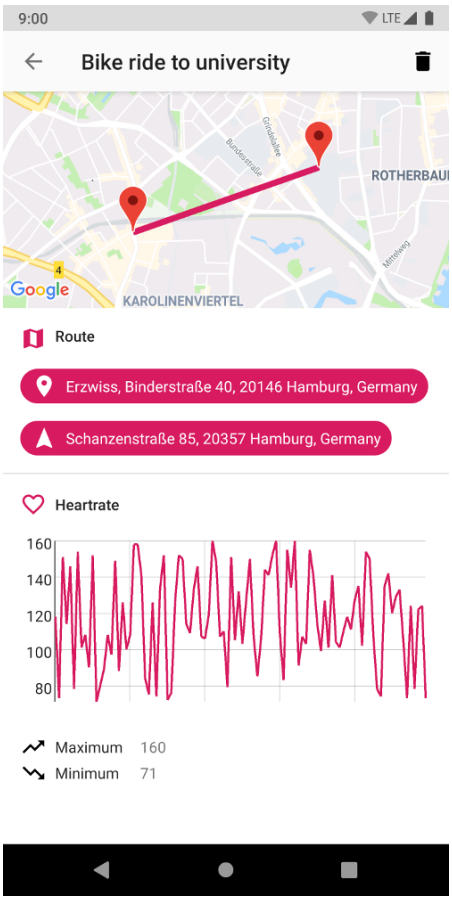


Figure 6.12: eHealth demo Data Activity.

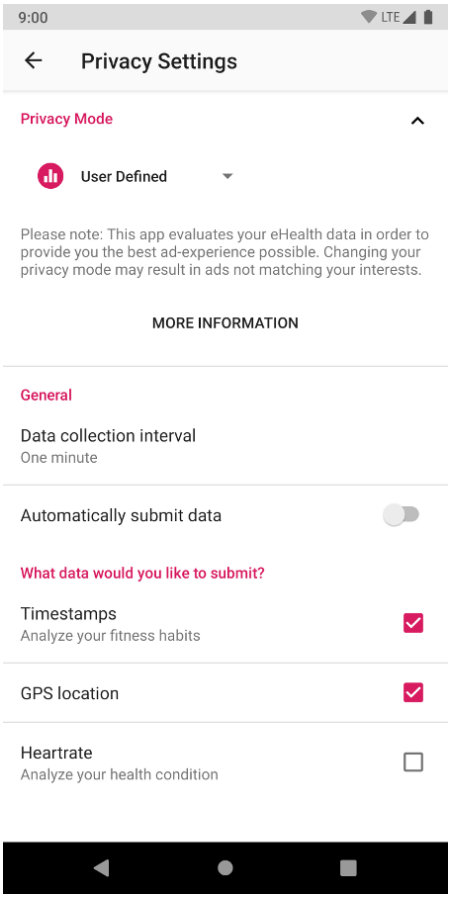


Figure 6.13: eHealth demo Privacy Settings.

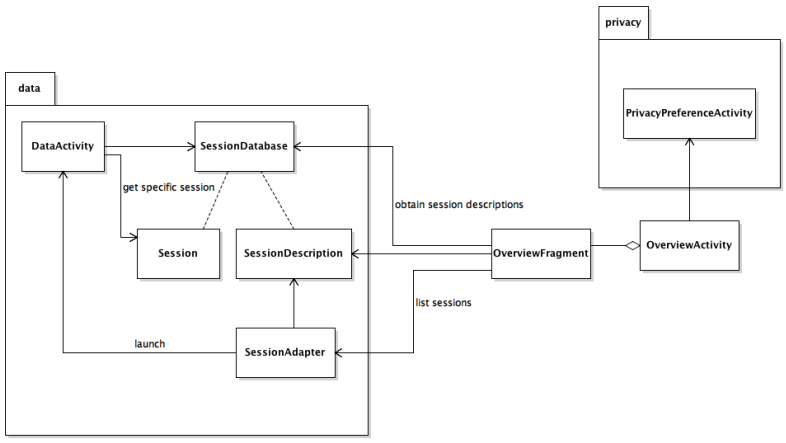


Figure 6.14: TRANK Privacy Data Flow.

via the "*User Defined*" Privacy Mode.

The main storage module of the application is an SQLite database containing all recorded session data together with the underlying Java classes representing the ses-

sion data, it makes up the core package of the app. The second component is the privacy package containing all privacy mode related classes. As figure 6.14 shows, the main `OverviewActivity` acts like a bridge between those elements and uses the `SessionDatabase` to obtain a list of all sessions available. The `SessionAdapter` class is then used to list them in different allocations and launch `DataActivity` instances when tapped. Similarly, a `PrivacyPreferenceActivity` instance is launched if the user wishes to change his privacy preferences. To the best of our knowledge, this is the first kind of implementation that considers the trade-off between privacy and system functionality. The ability to change through difference privacy modes without having to loose functionality indicates that TRANK can be used to model other applications that may require a high level of privacy protection by changing the privacy modes without collecting immense amounts of personal data from customers.

6.4.3 TRANK Based Database

In a real use case, the `SessionDatabase` would obtain its data from some kind of device sensors. However, due to this being a demo, the database is populated from JSON files containing mockup data. The API contains basic IO methods to list, add and remove sessions from the record. In the following, we present the application component design which contains a detailed description of objects used to implement the framework. It defines, packages, classes and methods used in the coding process.

Packages: The packages used to implement the eHealth prototype include the *Data* package, the *UI (User Interface)* package and the *Privacy* package. The ***healthdemo.data*** package is the core of the application and contains classes to access the underlying SQLite database and the actual data models. It also contains a number of related UI-widgets. The second package is the ***healthdemo.privacy*** package which contains all privacy related classes. Its core element is the `PrivacyMode` enum describing all available PrivacyModes. It also contains a number of related UI-widgets. The `healthdemo.ui` is a User Interface package which contains all Activities and Fragments making up the actual UI. Classes in this package use both the data and privacy packages. Within these packages we

implemented functionalities which include the `DataActivity`. This displays detailed information about a Session. The `MainActivity` provides the main interface of the app where the user can select different sets of data and access his current PrivacyMode settings. It consists of multiple fragments for compatibility with other applications. Measurements are done using a `Measurement` class. This is a data class representing a single fitness measurement consisting of different attributes such as a timestamp, location and vital functions. We implemented 5 most recent sets of driving data which are found in the `OverviewFragment` It provides an overview of the user's `PrivacyMode settings`. The privacy modes enable users to select permissions according to their preferences. These are implemented using the `PrivacyMode` class. **`PrivacyModeAdapter:`** The **`PrivacyModeAdapter`** takes Privacy Modes and displays them in an attached ListView or a Spinner. Adapters are used to populate ListViews with data. ListViews on the other hand are used to show a vertical list of scrollable items while Spinner objects are used to select one value from a set of items in a drop down menu.

`PrivacyModeChooserFragment:` The **`PrivacyModeChooserFragment`** lets the user choose his preferred PrivacyMode. Fragments are reusable parts of a User Interface which are embedded in activities. They are dependent on activities and cannot run on their own. The `PrivacyMode` class is an enumerated class which defines all available privacy modes as shown in the code snippet in Listing 1.

Listing 1: PrivacyMode

```
public enum PrivacyMode {
    UNKNOWN(-1, R.string.privacy_mode_unknown_title,
    OBFUSCATION(100, R.string.privacy_mode_obfuscation_title,
    R.drawable.ic_privacy_mode_obfuscation);
    public static PrivacyMode[] userModes() {
        List<PrivacyMode> modes = new ArrayList<>();
        for (PrivacyMode m : values()) {
            if (m.getID() >= 0) {
                modes.add(m);
            }
        }
        return modes.toArray(new PrivacyMode[modes.size()]);
    }
    public static PrivacyMode fromID(int id) {
        for (PrivacyMode mode : PrivacyMode.values()) {
            if (mode.getID() == id) {
                return mode;
            }
        }
                return UNKNOWN;
    }
 }
```

PrivacyModeChooserFragment.ModeChangedListener: The `ModeChangedListener` interface is used to notify an attached Activity about PrivacyMode changes. Custom Listeners are used to implement asynchronous event handling. They are also referred to as the Observer design pattern in which an object(subject) possesses observers which

inform the object automatically whenever the state of the object changes through method calls. Listeners can be used for example, to attach a click event to a button, I/O events and view events used in screens.

`PrivacyModeCustomizationFragment:` The `PrivacyModeCustomization` Fragment provides a SettingsFragment used to customize the "User Defined" privacy mode. The User Defined mode is a Privacy Mode used to select user defined data preferences to be used by the application.

`PrivacyModeView:` This View is used to display the `PrivacyMode` of the user. All available Privacy modes are listed in the PrivacyMode enum and can either be set programmatically via setMode(PrivacyMode) or in the XML file using the 'score' attribute. In case no mode was set the View displays a placeholder without any information.

`PrivacyPreferenceActivity:` The `PrivacyPreference` Activity contains both the PrivacyModeChooserFragment and PrivacyModeCustomizationFragment which together make up the privacy settings. An Activity is used to launch an app. It invokes call back methods that are used in specific stages of the app life cycle.

`Session:` This a data class representing a fitness session consisting of the `SessionDescription` attributes and a list of measurements.

`SessionAssetReader:` This class parses the mockup session assets (JSON) in '/assets/sessions/' and converts them into Session objects. It uses the JsonReader class provided by the Android SDK.

`SessionDatabase:` This class manages the SQlite database containing all recorded fitness Sessions. It includes methods to add, remove or list Sessions. By default it is populated with mockup data provided by the SessionAssetReader class. `SessionDescription:` This is a data class providing basic information about a Session such as it's ID and SessionType.

`SessionDescriptionAdapter:` This class is a subclass of BaseAdapter and functions as a List- or Spinner-Adapter for session data in form of SessionDescriptions. For each cache entry a view representing the basic session info will be generated.

`SessionType:` The SessionType is an enumeration class that defines all available types of fitness sessions.

SplashActivity: This Activity shows a splash screen on each start of the application. A splash screen is an introductory or welcome screen which consists of a logo, an image and the current version of the software. This offers the first experience of the application to the user.

TelematicsGraphView: This View is a more powerful GraphView featuring an additional title, headline as well as maximum and minimum value statistics.

TimelineFragment: The `Timeline` Fragment provides a list of all Session entries in a reverse (newest-to-oldest) order.

This chapter presents the evaluation of the prototype presented in chapter 6. The prototype is implemented as a TRANK based eHealth application and a TRANK based Telematics application. The evaluation was performed based on a Usability and user experience survey. Results of the survey were analyzed based on average results, a System Usability Scale scoring, and a Man-Whitney U test, after applying the TRANK framework integrated in the tested apps.

6.5 Privacy Evaluation Survey

To evaluate the privacy implementation in the apps, a survey was conducted with 25 adults recruited from Lancaster University. The aim of the survey was to determine how users perceive the privacy preserving features integrated in the eHealth Mobile application and the Telematics application after applying the privacy aware framework TRANK. The survey is a questionnaire based survey which was conducted based on the developed prototypes.

6.5.1 Materials and Methods

The questionnaire consists of three major parts; (1) general information about a participants gender, age and educational background, (2) information about the use of mobile apps and downloads and (3) information about the rating and privacy concerns regarding the privacy modes and features integrated in the app. Initially, the questionnaires were sent to participants through emails and their feedback transcribed accordingly. The

Table 6.1: Privacy Survey

	No participants
Gender	
Male	17
Female	8
Age	
18-20	2
21-25	11
26-30	4
31-40	5
45	3
Educational Background	
High school graduate	2
Undergraduate degree	13
Postgraduate degree	5
Doctorate	4
Prefer not to say	1
Use of Android apps	
Yes	19
No	6
Participants who liked the privacy features	
Yes	21
No	4
Participants concerned about data collection	
Yes	19
No	2
Did not care	4
Rating of the privacy apps	
Very Good	5
Good	18
Fair	2
Participants concerned about private information collected	
Very concerned	15
Concerned	6
May be concerned	4
Participants facing problems with privacy features in the apps	
Yes	2
No	17
Maybe	6

questionnaires were sent to approximately 70 participants with a link to the questionnaire using Google forms. The research has been approved by the Lancaster University Ethics committee. Prior to starting the study an information sheet and a consent form was sent to the participants. The information sheet is used to inform users about the study and its

contents. The consent form is used by participants to approve taking part in the study. Copies of the information sheet and the consent form can be found in the Appendix of the thesis.

6.5.2 Data analysis and metrics

We transcribed the data manually and extracted details using Microsoft excel. We analysed the data using IBM SPSS statistics software using descriptive statistics for the 10 questions in the questionnaire. A summary of the participant demographics is presented in Table 6.1.

Results

Participant response to the emails was fair at a response rate of 62%. Some responses though were not considered in the analysis as they did not meet the standards required. The results from the general part of the questionnaire based on gender, age, and educational background were as follows; Most of the respondents were male (68%, n = 17) compared to females (32%, n=8). Participants were mainly between 21 to 25 years of age (44%, n = 11) and the least number of participants was above 45 years (12%, n= 3). The highest level of education was that of a doctorate (16%, n=4) and the majority were undergraduates (52%, n=3).

In the second part of the survey, we emphasized on how participants perceived the privacy preserving features integrated in the apps. More that half percent of the participants downloaded Android apps (84%, n = 19) while 16% (n = 6) used other mobile operating systems. Of the 25 respondents, 21 participants (84%) liked the privacy features integrated in the app while 76% participants (n = 19) were concerned about the data collected in the apps, 8% (n = 2) respondents did not mind about data being collected in the apps and 16% (n = 4) did not care about data being collected. The apps were rated Good by most of the participants (72%, n=18) which was based on the privacy settings implemented in the apps. The majority of the participants were concerned about private information (PII) e.g. national identification numbers, credit card numbers to be collected by the app (60%, n=15). Finally, the majority of the participants did not find

any problems using the integrated privacy features (68%, n=17).

Discussion

In general, the key findings depicted that the participants liked the privacy features modes integrated in the apps, especially the possibility of choosing through different privacy modes when using the app. Users are able to choose from a low privacy mode to a higher privacy one without affecting functionality. This feature enables users to take full control of the type of data to be transferred to the app without constantly switching through a lot of privacy settings as it is in current modern apps. Instead a user can choose a privacy mode that contains the privacy settings they would like to use at default.

Limitations of the study

The major limitation of the study is that currently the apps are implemented in the Android mobile operating system only. The applications should be programmed in other mobile operating systems to determine the differences and how the features would be perceived by users.

Furthermore, the participant pool of 25 participants needs to be increased to get a broader interpretation of the features built in the app. This way more opinions about the apps can be gained and more ideas can be used to improve the apps.

6.6 Usability Evaluation Survey

We conducted a survey to investigate and evaluate the usability of the demo apps over a three week period. Determining the usability of the apps is a fundamental requirement in system development to assess user experience and the ease of use of a product after development. Usability is " the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use." [146]

There are five distinct usability components that are used to determine the quality of a product or system [147]. Among these we focused on three components in the app survey. These are *learnability*, *re-usability* and *system potential*. *Learnability* plays a major role in determining the cost a user requires to achieve a competent level of performance on a task

when using the app. We further focused on *re-usability* which is the performance level a user requires to perform a task when there is a continuous period of non-use. *System potential* determines the performance a user can obtain from the app.

Many researchers have used usability testing to investigate and identify the ease of using a product or system. "Usability testing refers to evaluating a product or service by testing it with a group of representative users" [148]. This approach is extremely valuable for identifying if the product is understood, easy to learn and easy to operate under a supervised period of time. In particular, a System Usability Scale (SUS) has been developed to evaluate industrial systems, and has been used to evaluate the prototypes in this study. A System Usability Scale is a ten-item Lickert scale which gives a general view of subjective assessments of usability [1].

Participants were recruited using the Security Lancaster mailing group and using friends. They were sent emails which included a participant information sheet with details of the study and a consent form requesting for their permission to take part in the study. The participant information sheet and the consent form can be found in the appendix of this thesis. After participants filled in the consent form and accepted to take part in the study, they were requested to rate the usability of the apps by submitting in an online Google form questionnaire.

Participants: Respondents were mainly from the academic domain at Security Lancaster, Lancaster university (N=50) and from Hamburg University (N=10). The population was provided by recruiting through emails which were mainly sent to full time researchers. This was done to provide a diverse participant base from the security and privacy backgrounds and to get a more technical base of current students who often use apps. Participants were given links to the apps and were requested to download them and study the privacy features integrated in the apps. They were asked to fill in the questionnaires based on the usability of the app.

Survey design: The survey presented participants with questions based on the System Usability Scale (SUS) used to measure the usability of interactive systems. An example of the original SUS questions is presented in table 6.2, rated from 1 to 5.

Table 6.2: System Usability Scale Statements [1]

	Strongly Agree		Strongly Disagree
I think that I would like to use this system frequently	1	-	5
I found this system unnecessarily complex	...		...
I thought this system was easy to use	...		...
I think that I would need the support of a technical person to use this system	...		...
I found the various functions in this system were well integrated	...		...
I thought there was too much inconsistency in this system	...		
I would imagine that most people would learn to use this system very quickly	...		
I found this system awkward to use	...		
I felt very confident using this system	1	-	5
I needed to learn a lot of things before I could get going with this system			

Table 6.3: Modified SUS questions used in the survey

	Average	SD
I think that I would like to use this app frequently	4.6	0.78
I thought that there was too much inconsistency in this app	1.4	0.52
I thought the app was easy to use	4.7	0.82
I found the system very cumbersome to use.	2.6	1.73
I thought there was a lot of consistency in using this app.+	3.8	0.78
I needed to learn a lot of things before I could get going with this app	1.8	0.56
I found the various functions of the app were well integrated	4.2	0.75
I think that I would need assistance to be able to use this app	2.8	0.91
I would imagine that most people would learn to use this systme very quickly	3.3	0.42
I think that I would need the support of a technical person to be able to use this app	2.4	1.35

The complete questionnaire consisted of questions based on a 5-point Likert scale which asked participants about the apps in the following order: (a) how frequent would they use the apps (b) the ease of using the apps (c) how consistent they found the privacy features integrated in the apps (d) the ease of technology used in apps. The questions were modified to match the original SUS questions as shown in table 6.3. A total of 75 emails were sent out to the Security Lancaster mailing list anticipating a response rate of 50%. The response rate was very low at 20% which led us to recruit new participants through other universities (Dresden University of Technology and Hamburg University). After checking for invalidity and inconsistencies, a sub sample of 50 participants was selected in the preliminary analysis. The process of recruiting participants was approved by the Lancaster University Faculty of Science and Technology ethics committee.

Measures:

The SUS Likert form consists of 10 questions in which the respondents rate the

Table 6.4: Sample Results obtained

ID	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Question 7	Question 8	Question 9	Question 10
001	Agree	Neutral	Agree	Disagree	Agree	Agree	Disagree	Agree	Disagree	Disagree
002	Neutral	Agree	Agree	Disagree	Agree	Neutral	Neutral	Agree	Neutral	Disagree
003	Disagree	Agree	Agree	Disagree	Agree	Neutral	Disagree	Agree	Neutral	Disagree
004	Strongly Agree	Agree	Agree	Neutral	Agree	Neutral	Disagree	Agree	Disagree	Disagree
005	Agree	Agree	Agree	Strongly Disagree	Agree	Agree	Disagree	Disagree	Agree	Disagree
006	Strongly Agree	Agree	Strongly Agree	Disagree	Agree	Agree	Strongly Disagree	Neutral	Disagree	Disagree
007	Agree	Agree	Strongly Agree	Disagree	Neutral	Agree	Disagree	Neutral	Disagree	Disagree
008	Neutral	Neutral	Neutral	Strongly Disagree	Disagree	Agree	Neutral	Disagree	Strongly Disagree	Strongly Disagree
.....										
050	Neutral	Agree	Strongly Agree	Strongly Disagree	Disagree	Agree	Neutral	Neutral	Strongly Disagree	Disagree

Table 6.5: Points assigned

Response	Points
Strongly agree	= 5 points
Agree	= 4 points
Neutral	= 3 points
Disagree	= 2 points
Strongly disagree	= 1 point

usability of the Ehealth app and Telematics app according to the integrated privacy features. Five features are positive (*1, 2, 3, 4, 5*) and the remaining five are negative (*5, 6, 7, 8, 9*). The response from both questionnaires were based on a 5-point Likert-type scale (*1=strongly agree, 2=agree, 3=Neutral, 4=disagree, 5=strongly disagree*) which respondents answered according to how they perceived the privacy measures integrated in apps. These SUS based questions were designed specifically for this study. The results of the study were then transcribed as shown in the Table 6.4.

6.6.1 Questionnaire analysis

To analyze the data collected from the questionnaires the samples were computed by assigning values to the each response to get a single average value. The values were assigned to the responses as shown in Table 6.5. To compute the values the responses were split into two parts. First we computed the results of the positive responses followed by the results of the negative responses for both the Telematics app and the eHealth app. To compute the final results we took a hypothetical approach based on calculating the percentages of the responses.

Table 6.6 illustrates how the points are assigned and computed. For example, given 20 participants who respond to a question with a "`Strongly Agree`" response, the number

Table 6.6: Calculated points

Responses		**Points**		
20 strongly agree	*	5 points	=	100
13 agree	*	4 points	=	52
10 neutral	*	3 points	=	30
5 disagree	*	2 points	=	10
2 strongly disagree	*	1 point	=	2
TOTAL				**194 points**

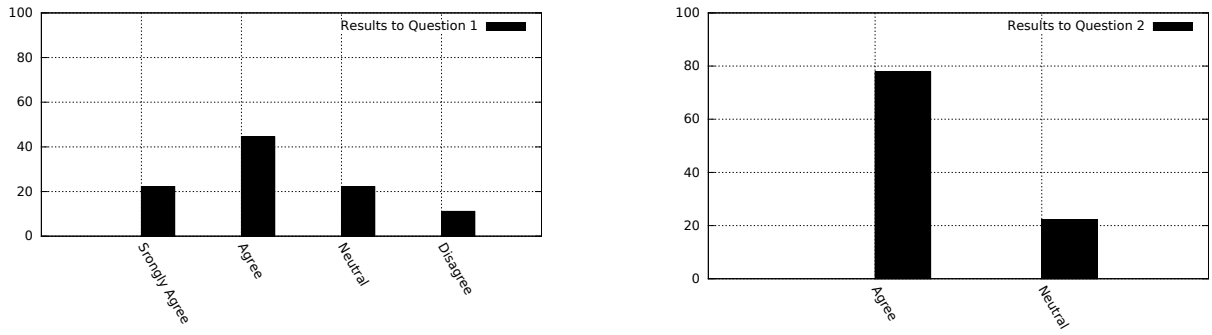
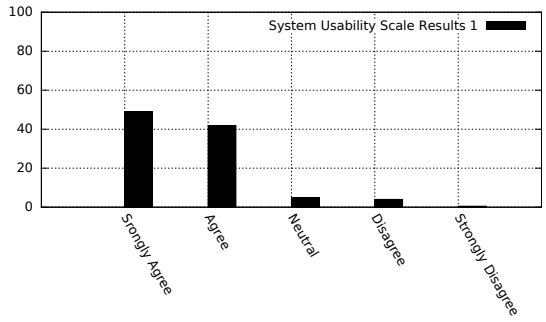


Figure 6.15: Response to Question 1 and 2

of responses value is multiplied with the assigned value given in Table 6.5. An example of the computation is shown in Table 6.6, given 50 participants who took part in the study, we take 194/50 = 3.88 as the average value for the Telematics app. The remaining values were further computed into averages.

Figure 6.15 shows the computed averages for *Question 1* which is; "I think that I would like to use this app frequently", and for Question 2 which is; "I thought that there was too much inconsistency in this app" respectively. All values were computed in this manner for all questions in Figures showing results 1 to 10 and for both demo apps (Telematics and Ehealth) to determine the average values.

eHealth app: Of the 75 participants who were sent emails to complete the survey, a total of 50 respondents were selected . Of these, 90% were male and 10% were female. Participant responses to the apps differed significantly. Although both links to download the apps were sent simultaneously, respondents were reluctant to complete the survey about the Telematics app in comparison to the eHealth app. Question one required the participants to respond to the following statement, "I think that I would like to use



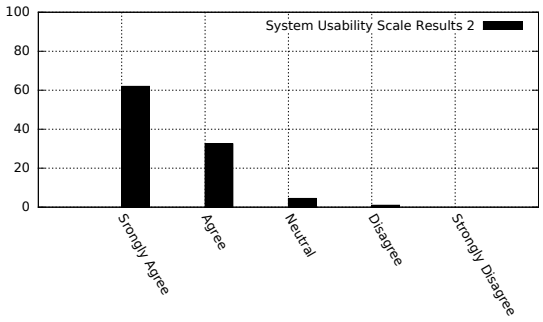
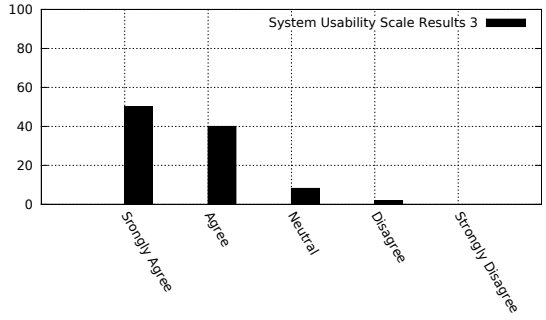


Figure 6.16: Result 1 and 2



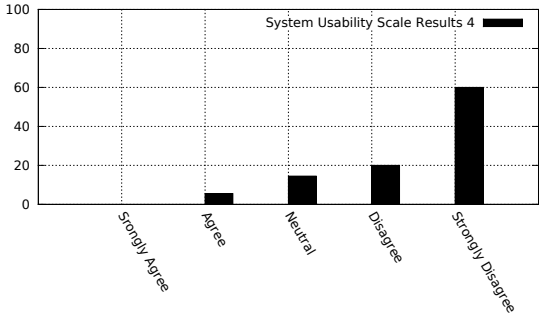
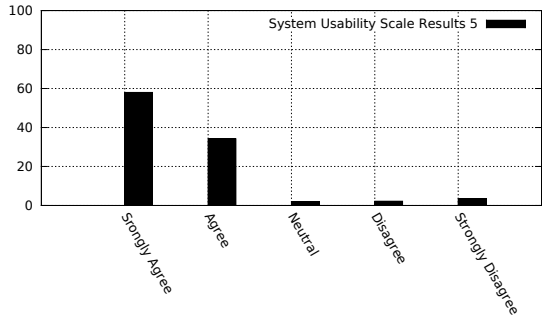


Figure 6.17: Result 3 and 4

`this app frequently.`" Participants were asked to rate the app using a 5 Likert-type scale anchored by `Strongly agree`/`Strongly disagree`. While statistical differences were not found between both apps, results did indicate that users were willing to use the eHealth fitness app more frequently, 44.4% responding with Agree, 22.2% Strongly agree, 22.2% Neutral, (t=-10.41, $p<0.05$). Analyzing the data to find out how users found its integrated privacy features revealed that the eHealth app was easy to use (77.8% Agree). In response to the consistency of using the app, it was found that 55.6% Agreed, 22.2% Strongly agreed and 22.2% were responded with Neutral. The statement" `I would imagine that most people would learn to use this system very quickly`" resulted into 66.6% of the participants responding with Agree and 33.3% with Neutral.

Responses from the Telematics app differed moderately to those of the eHealth app, in general participants are more familiar with using eHealth fitness apps compared to Telematics apps. To the statement "I thought the app was easy to use." 33.3% of the



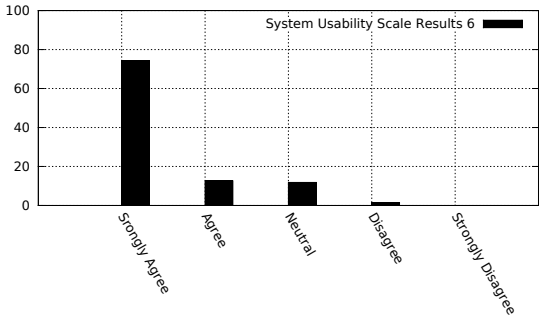
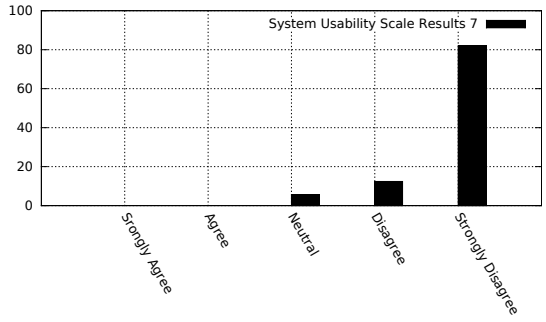


Figure 6.18: Result 5 and 6



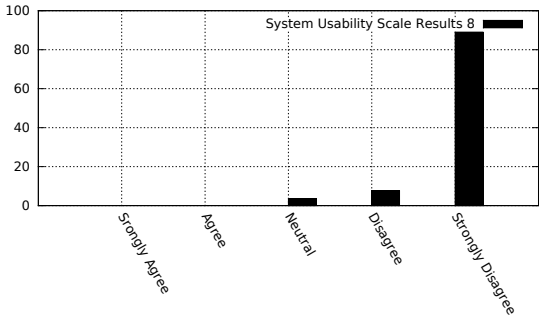
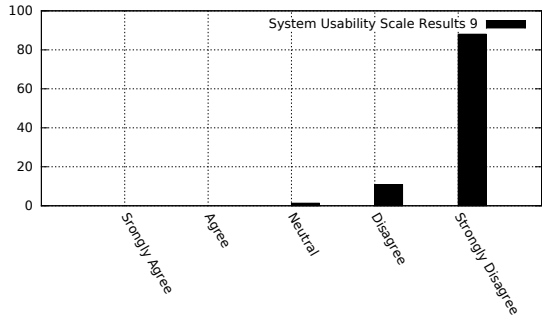


Figure 6.19: Result 7 and 8



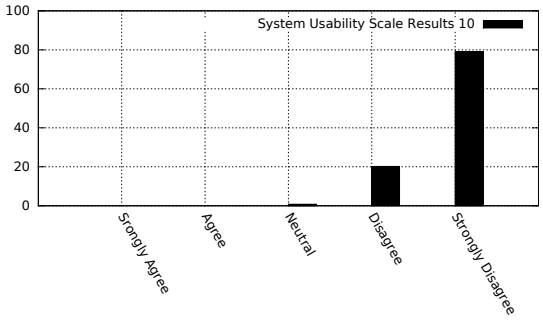


Figure 6.20: Result 9 and 10

participants responded with Strongly agree, 16.7% Neutral and 50% Agree. Participants found the apps consistency good and responded to the statement "`I thought there was a lot of consistency in using this app`" with 66.6% Agree and 33.3% Neutral. The apps ease of technical use was also found to be good with 66.7% respondents replying with Strongly disagree and 33.3% Disagree to the statement "I think that I would need the support of a technical person to be able to use this app".

6.6.2 System Usability Scale scoring

The results were further analyzed using the System Usability Scale (SUS) Score. We used the SUS scores because its a standard scoring system and its well established in measuring usability of industrial systems. The score is obtained by first placing the positive and negative items alternatively. That is the positive statements are situated in positions 1,3,5,7,9 and the negative statements in positions 2,4,6,8,10. To compute the final SUS score the values in the positions 1,3,5,7 and 9 have a score contribution of scale position minus 1. For positions 2,4,6,8 and 10, the score contribution is 5 minus the scale position. The overall SUS score is got by multiplying the sum of the obtained scores by 2.5. [149] and is summarized in [150] as follows;

- Step 1: Convert the scale into number for each of the 10 questions
  - Strongly Disagree: 1 point
  - Disagree: 2 points
  - Neutral: 3 points
  - Agree: 4 points
  - Strongly Agree: 5 points
- Step 2: Calculate
  - X = Sum of the points for all odd-numbered questions â€“ 5

Table 6.7: System Usability Scores

SUS Questions	Position	Value	SUS Score
I think that I would like to use this app frequently	5	5-1	4
I thought that there was too much inconsistency in this app	1	5-1	4
I thought the app was easy to use	5	5-1	4
I found the system very cumbersome to use.	1	5-1	3
I thought there was a lot of consistency in using this app.	4	4-1	3
I needed to learn a lot of things before I could get going with this app	1	5-1	4
I found the various functions of the app were well integrated	5	5-1	4
I think that I would need assistance to be able to use this app	4	4-1	3
I think that I would need the support of a technical person to be able to use this app	2	5-2	3
TOTAL			**80**

– Y = 25 - Sum of the points for all even-numbered questions

SUS Score = (X + Y) x 2.5

An illustration of the SUS scoring is shown in table 6.7 to give a final SUS score of 80. However, it is important to note that SUS scores which range from 0 - 100 are not percentages and should not be mistaken for percentage values. The scores are graded as follows; 51 points(F) and below are a Fail, 51-68 points(D) are Poor, 68 points(C) is Okay, 68 - 80.3 points(B) is Good and above 80.3 points (A) is Excellent. Understanding what the scores mean is explained by Bangor et. al [151] and in [150].

On the whole, products with a score below 51 are a cause of significant concern as this is considered as a fail. An overall score of 68.9 was obtained in the System Usability scale scores. A minimum of 50 was recorded and the maximum score was 92.5. According to the rating in [151], the mean score of 68.9 got in this study indicated that the apps were perceived to be good.

6.6.3 Assessment of User Behaviour attitudes

To further analyze the data, a Mann-Whitney U test [152],[153] was used to identify the behavioural attitude of the users towards the two apps and to find out how satisfied users were when using the apps. The Mann-Whitney U Test is also referred to as the Wilcoxon rank-sum test and is computed using the formula below:

Wilcoxon Rank-Sum Test

$$z = \frac{W - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}} \tag{6.1}$$

SUBJECT: this test is specifically designed to determine whether two *independent samples* came from equivalent populations. NOTE: this test is an alternative to the Two-Sample t test, however it does **not** require that the two populations follow the normal distribution and have equal population variances.

where:

- n_1 is the number of observations of the first population which is the eHealth app population in this thesis.
- n_2 is the number of observations of the second population which is the telematics app population in this thesis.
- W is the sum of the ranks from the first population.

The Mann-Whitney U test is a non-parametric test that is used to compare two sample means coming from the same population. Its used to test if the means are equal or not. It is widely used in fields like health care, psychology and business to compare attitudes or behavioural trends. For example, in business it can be used to find out the preferences of different people and see if these preferences change according to a given factor [154]. In this context, we carried out a Mann-Whitney U test to determine which app do the users prefer. We used an online Social science statistics calculator [155] to compare the median values of the responses to the SUS questions for both the eHealth app and the Telematics app. This was done to determine if there should be any changes in the design and implementation of the apps. We carried out the tests for the following positive questions;

1. I think that I would like to use this app frequently.
2. I thought the app was easy to use.

3. I thought there was a lot of consistency in using this app.

4. I found the various functions of the app were well integrated.

The null hypothesis is that there is no difference between the median of population 1 and the median of population 2.

H_0 = the null hypothesis is that the two medians are the same.

H_1 = the alternative hypothesis is that the two medians are not the same

We ran a Mann-Whitney U test to evaluate the differences in the responses of the Likert results reported in the apps. We found non significant effects of the two groups for all questions. For example, the Mann-Whitney U test result for question 1 was (Z score is -0.22024, the p-value is 0.82588), so the result is not significant at $p<0.05$. From the results obtained we have no evidence to reject the null hypothesis because the median for the two groups is the same. Therefore, from the analysis of the data reported, we conclude that the apps were equally perceived by the participants, and consequently, the design and implementation of the apps did not have to be changed.

6.6.4 Alternative usability analysis methods

Several methods have been proposed to measure usability through questionnaires. There are a number of alternatives to the SUS which can be used e.g. the Net Promoter Score, Usability Metric for User Experience (UMUX)[156], UMUX-LITE [157], SUPR-Q [158]. We chose to use the SUS as its well documented and easier to use. The methodology used in this study is the standard SUS questionnaire format. This questionnaire is based on the scientific analysis originally created by John Brooke [149]. However, approaches like the Net Promoter Score [159] are similarly widely used. The Net Promoter score consists of a questionnaire with one question and is used to measure customer experience.

Limitations: Limitations with the apps was depicted mainly in the new functionality that was integrated in the app which users were not familiar with. Response to the statement "`I found the various functions of the app were well integrated.`" resulted into 77.8% participants responding with Agree, 11.1% with Neutral and 11.1% with Disagree. The evidence suggests that some features were not well perceived by the

participants and need to be revised. Despite the high positive ratings of the apps, more effort has to be done in educating participants in how to perceive new approaches of integrating privacy in mobile applications. It can therefore, be inferred that users are not used to limiting the amount of data when using mobile apps. They openly expose their data to the apps in turn for better functionality. Limiting data exposure is seen as a burden especially in new technologies. Consequently, users are willing to give up their privacy in return for quicker and faster technologies.

General Overview: Overall, the usability of the developed apps was positively perceived by respondents to the study. The study met the design requirements that were positively assessed using the SUS scale and measured using the Likert scale. This study found high acceptance of the privacy features that were integrated in the eHealth fitness and Telematics apps. Therefore, the findings of this study suggest that users are capable of accepting new privacy implementations in future mobile applications once these are put on the market. In the following, we elaborate more on the key findings observed during the evaluation of the apps.

Privacy requirements: The evaluation of the apps shows that the privacy requirements we proposed have been met and are well perceived by users. The framework we designed that uses the methodology of privacy-by-design and privacy-by-default, to implement the trade off between functionality and privacy, improves the design of mobile applications as proposed in this thesis. Strong evidence has been given to emphasize that the privacy aware framework proposed has a high degree of accuracy when meeting privacy requirements, that its adaptable to real world scenarios as presented in the generated mobile applications, and that it is valid as we observed in the responses from the survey that we undertook. Therefore, the analysis given in this chapter has shown that the TRANK framework meets the objectives of the thesis that were put forward in Chapter 1 and Chapter 2.

Trust in new mobile applications applications: In the course of the studies we carried out, we observed that users have key trust issues in new mobile technologies. This was highly depicted in the choice of the apps they wanted to test. We noticed that users were reluctant to use the telematics app as this was considered very privacy infringing

compared to the eHealth privacy app. Once users are informed of the effects of sharing their data and location information, they are hesitant in using a mobile application. Therefore, mobile application developers should design easy to use applications that reduce the amount of PII collected. They should also inform users of the effects of the data that is collected. Currently applications on the market provide an On and Off solution to choose which features to be used without informing them of the exact impact of the data collected. In our this research we clearly, informed the users of the data collection methodologies used and what is used with the submitted data. This included, the location based information in telematics apps, the surveillance methods used as well as the type of billing used. This enabled the users to determine which app gathered more data and which did not. Much as the eHealth fitness app gathered data, users were willing to use it in comparison to the Telematics app.

6.7 Summary

This chapter presents the prototype implementation of the privacy aware trade-off analysis framework TRANK. The first application is a telematics application developed based on the TRANK framework. The second application is an eHealth fitness application developed using the TRANK framework. The implementation includes the following packages, the *User interface* (UI) package which contains the classes for running the user interface, the *Data* package which contains the classes that manage the data, and the *Privacy* package that contains classes for running the privacy related functionality. The Telematics prototype contains the following packages; *Data, Privacy, UI* which are similar to the ones in the eHealth prototype and a *Helper* package. The *Helper* package contains helper classes that manage the applicaton views. The chapter illustrates how we developed the telematics and eHealth prototypes which are based on the TRANK privacy aware trade-off analysis framework we introduced in chapter 5.

Furthermore, this chapter presents an evaluation of the prototype implementation of the TRANK framework. The TRANK framework was applied to two applications. The applications were evaluated using three approaches, (a)based on a privacy evaluation

survey (b) based on a usability survey of the developed apps and (c) the amount of data they collect and the privacy policies they use as implemented in the apps. Privacy-by-design mechanisms were integrated in the applications to generate privacy aware applications with less amounts of data collection and better privacy policy design as compared to apps provided in current App stores. Overall the results of the application implementation is positive. We conclude that TRANK is a suitable method for privacy preservation in mobile applications even though there is still room for improvement.

Chapter 7

Conclusions

The mobile application development domain is rapidly evolving. This thesis has demonstrated that the integration of privacy preserving methodologies over the past years has improved and indications of new privacy enhancing methodologies are being developed. However, the increase of privacy breaches and data leakages suggests that there is need for enhanced privacy preserving mechanisms to reduce privacy invasion in future mobile applications. More than one half of the mobile applications investigated do collect data that is not required for system functionality and is privacy invasive. The studies carried out in this research make this observation as per the data collected and the analysis performed in this thesis. This observation is driven by the existence of third party service provider companies which aggressively collect massive amounts of data for advertising purposes.

Current mobile applications have integrated privacy policies as a means of integrating privacy in mobile apps, however, existing privacy policies come with a variety of limitations. They are not easily understood by customers and the majority of privacy policies are long and tedious to read such that users do not read them as required. In addition, the privacy policies are not in sync with the actual data that is collected by the apps both in the companies sites and within the apps.

Prior research has focused on encryption methods in an attempt to secure mobile applications. Although some research has emphasized on the need to protect user privacy,

little has been done to implement these ideas. Therefore there remains a gap to integrate privacy-by-design methodologies at the initial stages of product development. The tech giants e.g. Google and Facebook are the main funnels of data collection and we have seen reports about privacy invasion both from users and the data collecting companies- This phenomenon does not only increase mistrust in users but calls for proper privacy preserving solutions that ensure that user personal identifiable information is protected. This can be implemented not only in technical terms but with the introduction of legal structures that generate binding laws governing user privacy and that regulate the industry with a stronger upper-hand. Current regulations are either not suitable for the rapidly growing mobile application industry or are not enforced strongly leaving companies an option of massive data collection of private information and thus the increase of privacy abuse.

There is a sense that users are increasingly getting concerned about their information being abused by data collecting companies. This has been widely reported in the media when companies used this data to manipulate elections. Collaboration of both data collecting companies, users and privacy engineers is therefore required to ensure that a collective approach to protect user privacy is obtained. Otherwise users will have to live with the fears of their private information collected by tech giants and mobile apps especially with sensitive data like medical data of being compromised and exposed. Large companies like Google are keen in improving privacy protection in their applications but smaller app developers and mobile applications that have a limited budget but continue collecting sensitive data will be prone to future privacy attacks and data breaches. Never the less the willingness to improve user privacy in future mobile applications is increasing as users are gradually concerned about their private identifiable information being leaked to other companies like insurance companies or employment companies. Therefore, the work introduced in this thesis which introduces a privacy aware trade-off analysis framework plays a major role in the development of privacy preservation methodologies in future mobile applications and acts as a stepping stone in privacy research. The study would be furthered by implementing the framework and the framework based prototypes

using a different mobile application operating system to investigate the impact it has in another experimental setup.

This chapter concludes by revisiting the aims and objectives of the thesis presented in chapter 1. We analyze and evaluate to what extent the aims and objectives have been met. The purpose of this work is to enhance the integration of privacy preserving methods in future mobile applications. To achieve this goal, this thesis presents a novel privacy aware trade-off analysis framework (TRANK) which integrates privacy-by-design concepts in mobile applications while considering the trade-off between privacy and functionality. The aims and objectives were achieved in the following ways:

- Chapter 1 provides user awareness about data being collected by current mobile applications and highlights some of the data breaches that have occured in recent years.

- Chapter 3 reports on the surveys we performed to show the challenges of privacy implementation in V2X mobile applications.

- Chapter 5 illustrates how the privacy trade-off analysis is generated and implemented on two applications; the Telematics application and an eHealth fitness application. The objective met here is to support application designers in decision making on how to integrate privacy at the initial stages of application development at the same time considering the trade-offs between privacy and functionality.

- Chapter 6 presents how the framework is implemented and highlights on the improvement in privacy preservation gained by using the framework to design privacy aware mobile applications. It further presents an evaluation of the prototype implementation of the TRANK framework. The prototypes were evaluated based on average results, a Privacy user study, a System Usability Scale scoring, and a Man-Whitney U test to determine how users perceived the developed apps.

7.1 Contributions of this thesis

This thesis made minor and major contributions to the preservation of privacy in mobile applications in using the following procedures:

A. ***Major contributions:***

- We developed a privacy aware Trade-off Analysis Framework (TRANK) that ensures that privacy-by-design principles are integrated in the initial stages of application development. We identified that when designing privacy aware applications, the trade-off between privacy, functionality and performance needs to be addressed. We developed guidelines on how to implement the framework in other domains using the Telematics and eHealth applications.
- We developed guidelines for resolving privacy and functionality trade-offs including suggestions of using a trade-off matrix.
- We developed an implementation prototype for the Trade-off Analysis Framework to serve as a proof of concept for TRANK. Data Collection Scores and Privacy Policy Scores used to evaluate the framework were significantly improved by more that 70%.

B. ***Minor contributions:***

- This thesis has highlighted current data breaches both in the media and literature to sensitize users about the risks involved when sharing Personally Identifiable Information (PII) to data controllers and processing companies.
- We investigated privacy implementations in eHealth applications and V2X telematics applications to identify gaps in designing privacy aware mobile applications through an exploratory study.

In particular, the main contribution of this thesis is the privacy Aware Trade-off Analysis Framework TRANK. The framework does the following:

- Defines privacy goals of the system under development to determine which goals are important for system functionality and how they interplay with other goals under consideration.
- Maintains and manages privacy goals and requirements both at the mobile application level and the privacy policy level.
- Defines privacy, functional and performance requirements and how these interfere with each other to enhance privacy protection at large using privacy policies and enforcing GDPR compliance.
- Leverages the privacy requirements and functionality by reducing the amount of personal data collected without negatively affecting app functionality.
- Evaluates the privacy-by-design principles developed within the envisaged design framework, TRANK and in order to present the generic design principles, the prototype further evaluates over representative application domains that are prone to privacy breaches.

In the following sections, we elaborate on the key concepts of privacy preservation in relation to data collection, privacy policies and privacy regulation to mitigate the risks of designing privacy aware applications as observed in this research.

7.2 Data collection in mobile applications

Data collection in mobile applications plays a great role in designing privacy aware applications. Current mobile applications collect data to enable them perform according to the desired functionality. However, according to studies we made, data processing companies accumulate an immense amount of data that is not required for system functionality without users consent. Users have significantly lost any control over the data given to mobile companies through the services they provide. Mobile application services that currently gather a lot of data include; social networking services, web services and business applications. Customers provide personal information to such companies in

return for the functionality they provide. It is estimated that 50% of users are willing to provide their purchasing history e.g. when buying products online for any discounts offered. Although data breaches are often reported, users continue to submit their personal data in exchange for the services provided.

In this thesis, we therefore, call for a "minimal data collection" approach from application developers. App developers must request for only the data they need for app functionality. We also call for app developers to increase user awareness about the data collected. Application developer companies should inform users explicitly how they collect the data and for which purposes its collected for. Recognizing that minimal data collection is an important way of mitigating the risks of the data being compromised, consumers need guidance on how to achieve this goal and improve privacy in mobile applications. Currently most users have adapted to controlling with whom their data is shared especially in OSNs where they are given choices e.g., sharing with family, friends and to the general public. This option, however, has not been implemented in mobile apps. The data collected by app developer companies is solely controlled by the companies. They decide with whom to share the data e.g. with third party service providers. The user therefore has no control of their data.

We recommend app developers to allow for user empowerment over collected data so that users can be able to opt-in and out of the application. There should be measures that enable users to anonymize their data in cases where data sets are required for app functionality. The idea of opting in-and-out of the data should not, however, lead to application failure. Therefore, privacy engineers, software engineers and regulators should enforce minimal data collection as well as users ability to opt out of applications without losing functionality. This way the amount of data collected and eventually that is compromised in case of any imminent attacks is minimized.

7.3 Privacy policies in mobile applications

Privacy policies are designed with the aim of disclosing how personal information in mobile applications is used. All apps must include privacy policies in the Apple store

and Google Play store. The disclosures in privacy policies provide transparency on how the data is to be used. In this study, we found that mobile applications have integrated privacy policies on company websites and Google Play Store. However, most apps do not employ privacy policies during app installation. Some apps do have in-app privacy policies but they are not easily readable. We therefore, call for a better design of in-app privacy policies to enable users understand why companies collect data and how they use it. Furthermore, we observed that the privacy policies are lengthy and not readable on small portable devices like hand held phones. The highly technological terms in the policies make it very hard for a lay user to understand. We observed that users only tick policies presented to them without understanding what they are committing to when they check the boxes with "Yes" or "No". We therefore, call upon app designers to find means of simplifying the privacy policy design to enable users fully understand the terms given in the privacy policies and how they relate to their personal data.

7.4 Privacy regulation in mobile applications

We have seen an improvement in privacy regulation in the course of our study. In May 2018, the GDPR was enforced. This has led to companies integrating GDPR privacy regulations to protect user privacy. Furthermore, California introduced the California Online Privacy Protection Act (CALLOPA) law which ensures that application designers in California integrate privacy policies in system design. We, however, observed that, few mobile apps are GDPR compliant. Much as these laws have been passed, major online giants like Google and Facebook have not fully implemented them in their systems [57]. More regulation control should be enforced to ensure that mobile applications are compliant to the existing enacted laws and regulations. The major problem faced in enforcement of the law is that the regulators are often reactive and not preventive. This means that only when, for example, a massive data breach has occurred do the regulators respond. We therefore, call for regulators to be more proactive in privacy law enforcement.

7.5 Reflections on the research undertaken

This thesis presents the following key outcomes based on the research performed and the empirical studies undertaken.

7.5.1 The Privacy aware Trade-off analysis framework (TRANK)

The thesis initially presents the TRANK framework, which is a privacy-by-design methodology that aids system developers, privacy engineers ad system analysts to integrate privacy at the initial stages of system development putting into consideration the trade-offs that occur between privacy, system functionality and performance. TRANK is a design framework. To integrate privacy in the system development life-cycle we initially define the privacy goals the system has to fulfil. This can be achieved on a high level as required by the requirements process or detailed based on the architecture to be implemented. Based on the generated goals a privacy and functionality goals trade-off analysis is developed. This is the phase which initially introduces the aspect of trade-offs. Trade-offs are integrated at the first stages of the framework to enable a consistent approach in designing the privacy aware application. This is mainly because if the trade-offs are integrated at a later stage, more work has to be implemented in the development phase. Therefore, it is important to consider the trade-offs prior to reaching the system development stage. For each of the generated goals a trade-off analysis is performed, which is then used to generate the privacy requirements. The elicited privacy requirements are used to generate the privacy and functionality trade-off matrix. The matrix is developed based on a privacy trade-off analysis template. Finally the generated requirements are managed using a requirements management tool. This step is crucial in the mobile application software development phase. Due to the rapid development of mobile applications, privacy requirements keep evolving and thus need to be managed whenever new technologies are developed. These five steps are the core development aspects of TRANK. Once these steps are implemented in the requirements development phase, then an enhanced privacy aware application can be achieved that puts into consideration the privacy and functionality trade-offs during the initial mobile application development.

7.5.2 Empirical evaluation of the TRANK based eHealth and Telematics prototypes.

The evaluation of TRANK was done by performing two major surveys. A privacy evaluation survey and a Usability evaluation survey. The surveys were undertaken using the developed TRANK based prototypes. TRANK is a design methodology and complex to fully implement automatically. We therefore, used the survey approach in order to get an insight on how users perceived the applications both from a privacy perspective and a usability perspective. The surveys were used to determine the privacy concerns of users while using the apps especially with data collection and data sharing of sensitive private information entailed in PII. Another aspect that users were asked was how they perceived the privacy features that were newly integrated in the apps which enabled users to have control of the data being transferred to the app for further processing. The usability study on the other hand mainly focused on how users perceived the app based on 10 key questions used to determine the usability and quality of a developed system or product. The questions were based on three major usability components which are learnability, re-usability and system potential. Overall the TRANK based prototypes performed well. The participants perceived the apps positively and the general feedback on the usability and ease of use of the app reported from the participants was positive.

7.6 Research implications for emerging privacy-aware mobile applications

We recognise that designing privacy aware mobile applications is very important to mitigate privacy breaches and minimise current privacy abuse in mobile applications. This research has pointed out privacy concerns as reported in chapter 1. In an attempt to address these challenges, this research has proposed the use of a privacy-by-design framework that integrates privacy-aware design in future mobile applications using the proposed privacy aware trade-off analysis frame work TRANK. Based on the researched literature that has been presented in chapter 2, we emphasize that the use of TRANK as

a design methodology greatly improves privacy preservation in future mobile application development.

7.6.1 User sensitization

Sensitization of users about the risks involved in sharing private information is required especially by the app developers. Although most of the apps do have a privacy policy during app download, most of them do not have in-app privacy policies. I addition to integrating in-app policies in mobile applications, other forms of user sensitization can be implemented in the form of nudging users to use privacy related features when downloading the apps. The privacy settings should be integrated as the default settings in mobile applications, from here the users can then decide on which other features to use. Another way of nudging users to use privacy related features is to include privacy permissions as a default set up. Android permissions have been redesigned to inform users about which permissions to use, however most users do not understand what the permissions mean and so do not know how to use them. This means that more sensitization has to be done in an effort to improve privacy in future mobile applications.

7.7 Future Work

This section discusses future directions in which the privacy of mobile applications can be improved on the basis of the research findings of this thesis.

7.7.1 TRANK based design for other domains

One of the possible future directions is to integrate TRANK in other domains e.g. Finance, Manufacturing, Healthcare, eCommerce, Entertainment, to determine the privacy violations through the amount of data collection that these applications collect and address them using the TRANK framework. Detecting and identifying data collection in different domains enables privacy engineers determine which functionality require more data collection than others. In identifying the functionality of different domains, app designers are able to reduce the amount of data collected based on the chosen functionality. The

core aim of this approach is to identify which domain collects immense amounts of data in an effort to regulate it. There have been reports about massive data collection especially in domains like Online Social Networks [160] and the lack of trust towards user privacy but consumers continue to use these products. While it was shown that the TRANK framework improved user perception of privacy and privacy integration in the two mobile applications studied in this thesis, it is hypothesized that a detailed study on integrating TRANK in other domains will further contribute to the improvement of privacy preservation in these domains by producing better privacy-aware applications. Considering different domains as Financial mobile applications, OSNs or Online eCommerce domains like Amazon is important not only to the consumers but also privacy engineering researchers and developers in an attempt to improve privacy integration in future mobile applications.

7.7.2 Tool support for the trade-off analysis framework

The privacy trade-off analysis framework introduced in chapter 5 is manual. Application designers and privacy-by-design engineers will need to create the steps involved and generate the framework carefully and gradually. The privacy and functionality matrix needs to be clearly and thoroughly automated. An automated approach for designing the framework is beneficial to aid in implementing the framework from a privacy modeling design perspective. This can be achieved by automatically generating the privacy, performance and functionality matrix based on the requirements of the mobile application. This automated design can then be used in subsequent steps of the the framework to create applications that are easily and quickly implemented.

7.7.3 Regulated privacy-by-design templates

There is a need for privacy engineers to develop applications based on regulated app design templates that are uniform to reduce unnecessary data collection. A potential future development, therefore, is to design mobile apps using easily understood privacy preserving templates that emphasize privacy implementation and minimal data collection. Current research efforts have focused on privacy policy templates that aid application

designers to generate privacy policies [161], however, these templates are legally focused and not technically focused in terms of privacy enforcement. We observe that although the privacy policies generated are legally binding, they are no privacy enhancing technologies to integrate privacy-by-design concepts in mobile application development. Therefore, there is a need for designing templates that offer both the privacy-by-design concepts incorporated with the legally binding aspects of privacy preservation which can be utilized during mobile application development.

7.7.4 Data Collection Analysis

Big data analytics is able to examine large data sets from various sources to discover trends in marketing and predictive analysis. We are all aware that currently, large data mobile app controllers and processors (e.g. Facebook, Messenger, Whatsup, Youtube) use big data like Facebook Amazonsolutions for analyzing future market trends and uncovering hidden patterns and correlations about the products we buy or the services they offer [162]. A potential future technological development in privacy data analysis, therefore, would be to use big data technologies to analyze data collection trends especially with the new mobile applications on the market. We observer that data processors are increasingly collecting PII, health related data, marketing data with the promise of improving service through, for example, targeted advertising, however it would be worth investigating from a reverse engineering perspective which models are used to generate these adverts in an attempt to educate users on protecting their data. This way customers are cautious about the data they are willing to provide and which privacy preserving measures they need to undertake when using future mobile applications.

7.7.5 Regulated privacy enforcement in application development lifecycle

With the enactment of the GDPR regulation which governs personal data processing, companies which have been penalized for privacy and data breaches have been minimal. We therefore,like Facebook Amazon call for a more stringent privacy enforcement strategy that puts in place strict consequences to data controllers and processors when user privacy

has been compromised. At present, the legal framework presented through the GDPR regulations and other legal governing bodies are not being thoroughly implemented by companies. It is reported that six months after the GDPR was applicable big corporations like Facebook, Amazon and Google privacy policies' were not compliant to GDPR regulations [163]. As more private data published to large corporations, a further area for investigation is to enhance privacy enforcement during application development of mobile application.

7.7.6 Reflections and closing remarks

This thesis has introduced a privacy preserving Trade-off analysis framework and empirically evaluated it using a prototypical methodology and through privacy and usability surveys based on our findings. Using both qualitative and quantitative research methodologies, the results obtained are promising especially in exhibiting the importance of privacy preservation in future mobile applications. This observation goes beyond the aspects that have been assessed and calls for future enhanced methods of integrating privacy in the development process of mobile apps.

Users should be encouraged to take privacy preservation in a systematic way based on the type of data they would like to make public to the tech giants and that they should keep private. While existing privacy preserving frameworks support privacy protection, this research has added value to the privacy knowledge base especially when integrating privacy using the privacy-by-design notion. This is required especially to improve and support of the evolving privacy concerns. Therefore, future work should put an emphasis on how privacy-by-design and especially privacy-by-default should be integrated in all mobile applications.

Bibliography

[1] John Brooke. Sus: A quick and dirty usability scale. *Usability Evaluation in Industry*, 1996. [Cited on pages xii, 163, and 164]

[2] New York Times. Verizon will pay 350 million dollars less for yahoo. *https://www.nytimes.com*, 2017. [Cited on page 4]

[3] Taylor Armerding. The 18 biggest data breaches of the 21st century. *CSO*, 2018. [Cited on page 4]

[4] The Guardian. 50 million facebook profiles harvested for cambridge analytica in major data breach, 2018. [Cited on page 4]

[5] Dave Lee. Facebook security breach: Up to 50m accounts attacked, 2018. [Cited on page 4]

[6] Jr. John P. Mello. Healthcare security $65 billion market. *https://cybersecurityventures.com/healthcare-cybersecurity-report-2017/*, 2017. [Cited on page 5]

[7] Jessica Davis. Molina healthcare breached, exposed patient data for over a month, 2017. [Cited on page 5]

[8] Esecurityplanet. 21st century oncology notifies 2.2 million patients of data breach, 2017. [Cited on page 5]

[9] Joshua Berlinger. Singapore hack affects 1.5 million – including prime minister, 2018. [Cited on page 5]

[10] Pierluigi Paganini. Hackers broke into the celeb london bridge plastic surgery clinic, 2017. [Cited on page 6]

[11] Brian Moylan. Grindr was a safe space for gay men. its hiv status leak betrayed us, 2018. [Cited on page 7]

[12] Nivedita Balu Jim Finkle. Under armour says 150 million myfitnesspal accounts breached, 2018. [Cited on page 7]

[13] David Lumb. Fitness app pumpup left users' personal data exposed on server, 2018. [Cited on page 7]

[14] Alex Hern. Fitness tracking app strava gives away location of secret us army bases. *The Guardian*, 28, 2018. [Cited on pages 7 and 76]

[15] Amanda Jackson. Husband and wife never expected their fitbit would tell them this., 2016. [Cited on page 7]

[16] Peter Pitts. The privacy delusions of genetic testing., 2018. [Cited on page 8]

[17] Equifax. The dark web explained - what does it mean for on line security., 2018. [Cited on page 9]

[18] Olivia Solon. Credit firm equifax says 143m americans' social security numbers exposed in hack, 2017. [Cited on page 9]

[19] Ms. Smith. 2 canadian banks hacked, 90,000 customers' data stolen, 2018. [Cited on page 9]

[20] Carole Cadwalladr and E Graham-Harrison. The cambridge analytica files. *I made Steve Bannon's psyhological warfare tool': meet the data war whistleblower*, 2018. [Cited on pages 11 and 111]

[21] William Enck, Peter Gilbert, Seungyeop Han, Vasant Tendulkar, Byung-Gon Chun, Landon P Cox, Jaeyeon Jung, Patrick McDaniel, and Anmol N Sheth. Taintdroid: an information-flow tracking system for realtime privacy monitoring on smartphones. *ACM Transactions on Computer Systems (TOCS)*, 32(2):5, 2014. [Cited on pages 11 and 114]

[22] Li Li, Alexandre Bartel, Tegawendé F Bissyandé, Jacques Klein, Yves Le Traon, Steven Arzt, Siegfried Rasthofer, Eric Bodden, Damien Octeau, and Patrick McDaniel. Iccta: Detecting inter-component privacy leaks in android apps. In *Proceedings of the 37th International Conference on Software Engineering-Volume 1*, pages 280–291. IEEE Press, 2015. [Cited on pages 11 and 114]

[23] Alan F Westin. Privacy and freedom. *Washington and Lee Law Review*, 25(1):166, 1968. [Cited on page 14]

[24] Xiaodong Lin, Xiaoting Sun, Pin-Han Ho, and Xuemin Shen. Gsis: A secure and privacy-preserving protocol for vehicular communications. *IEEE Transactions on vehicular technology*, 56(6):3442–3456, 2007. [Cited on page 14]

[25] Fei Chen, Bezawada Bruhadeshwar, and Alex X Liu. A cross-domain privacy-preserving protocol for cooperative firewall optimization. In *INFOCOM, 2011 Proceedings IEEE*, pages 2903–2911. IEEE, 2011. [Cited on page 14]

[26] Nils Ulltveit-Moe and Vladimir Oleshchuk. Two tiered privacy enhanced intrusion detection system architecture. In *Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, 2009. IDAACS 2009. IEEE International Workshop on*, pages 8–14. IEEE, 2009. [Cited on page 14]

[27] Sarah Spiekermann and Lorrie Faith Cranor. Engineering privacy. *IEEE Transactions on software engineering*, 35(1):67–82, 2009. [Cited on pages 14, 17, 19, 24, 28, 29, 30, 43, and 146]

[28] GW Van Blarkom, JJ Borking, and JGE Olk. Handbook of privacy and privacy-enhancing technologies. *Privacy Incorporated Software Agent (PISA) Consortium, The Hague*, 198, 2003. [Cited on pages 14 and 15]

[29] LC Williams. The 9 biggest privacy and security breaches that rocked 2013. *Thinkprocess, December*, 31, 2013. [Cited on page 15]

[30] A. Dardenne, A. Lamsweerde, and S. Fickas. Goal-directed requirements acquisition. *Science of computer programming*, 1993. [Cited on pages 15 and 123]

[31] Yu Eric. Agent orientation as a modelling paradigm. *Wirtschaftsinformatik*, 2001. [Cited on page 15]

[32] R. Chitchyan, A. Rashid, P. Sawyer, A. Garcia, M. Alarcon, J. Bakker, B. Tekinerdogan, S. Clarke, and A. Jackson. Survey of aspect-oriented analysis and design approaches. In *Survey of aspect-oriented analysis and design approaches.Tech. rep. D11 AOSD-Europe-ULANC-9, AOSD-Europe*, 2005. [Cited on page 15]

[33] M. Deng, K. Wuyts, R. Scandariato, B. Preneel, and W. Joosen. A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineering*, 2011. [Cited on pages 15, 62, 69, 72, 77, and 118]

[34] K. Christos, E. Kavakli, and S. Gritzalis. Addressing privacy requirements in system design: the PriS method. *Requirements Engineering*, 2008. [Cited on page 15]

[35] OECD. Oecd guidelines on the protection of privacy and transborder flows of personal data. 2019. [Cited on page 15]

[36] Automotive Special Interest Group. Automotive SPICE. 2017. [Cited on page 16]

[37] Mockplus. Mockplus: Prototype faster. smarter. easier, 2018. [Cited on page 16]

[38] Balsamiq. Balsamiq rapid wireframing tool, 2018. [Cited on page 16]

[39] R. Kazman, M. Klein, and P. Clements. Method for Architecture Evaluation. *SEI Technical Report*, 2000. [Cited on page 16]

[40] S.H Houmb, G. Georg, J. Juerjens, and R. France. An integrated Security Verification and Security Solution Design Trade-off Analysis. *In integrated Security and Software Engineering: Advances and future visions*, 2007. [Cited on page 16]

[41] Ann Cavoukian and Michelle Chibba. Start with privacy by design in all big data applications." guide to big data applications. *Guide to Big Data Applications. Springer, Cham 29-48*, 2018. [Cited on page 17]

[42] Ann Cavoukian. Privacy by design. *IEEE Technology and Society Magazine 31.4, 18-19*, 2012. [Cited on page 17]

[43] Isaac Clarke. Gapp privacy : 10 generally accepted privacy principles, 2017. [Cited on page 18]

[44] J.Petit, F.Schaub, M.Feiri, and F.Kargl. Pseudonym schemes in vehicular networks: A survey. *IEEE Communications Surveys and Tutorials*, 2014. [Cited on page 19]

[45] M. P. Feiri, J. Y. Petit, and F. Kargl. Real world privacy expectations in vanets. *Proceedings of the 2nd GI/ITG KuVS*, 2014. [Cited on pages 19 and 22]

[46] E. Fonseca, A. Festag, R. Baldessari, and R. Aguiar. Support of anonymity in vanets - putting pseudonymity into practice. *IEEE Wireless Communications and Networking Conference (WCNC)*, 2007. [Cited on page 20]

[47] D. Foerster, F. Kargl, and H. Loehr. Puca: A pseudonym scheme with user-controlled anonymity for vehicular ad-hoc networks (vanet. *IEEE Vehicular Networking Conference (VNC)*, 2014. [Cited on page 20]

[48] F. Armknecht, A. Festag, D. Westhoff, and K. Zeng. Cross-layer privacy enhancement and non-repudiation in vehicular communication. *Communication in Distributed Systems*, 2007. [Cited on page 20]

[49] P.P Tsang and S.W Smith. PPAA: Peer-to-peer anonymous authentication. *International Conference on Applied Cryptography and Network Security*, 2008. [Cited on page 21]

[50] Y. Sun, R. Lu, X. Lin, X. S. Shen, and J. Su. An efficient pseudonymous authentication scheme with strong privacy preservation for vehicular communications. *IEEE Transaction on Vehicular Technology*, 2010. [Cited on page 21]

[51] A. Weimerskirch. V2X Security and Privacy: The Current State and Its Future. *ITS World Congress, Orlando, FL*, 2011. [Cited on page 21]

[52] B. Wiedersheim, Z. Ma, F. Kargl, and P. Papadimitratos. Privacy in inter-vehicular networks: Why simple pseudonym change is not enough. *International Conference Wireless On-demand Network Systems and Services*, 2010. [Cited on page 21]

[53] Rongxing Lu, Xiaodong Lin, Tom H Luan, Xiaohui Liang, and Xuemin Shen. Pseudonym changing at social spots: An effective strategy for location privacy in vanets. *IEEE transactions on vehicular technology*, 61(1):86, 2012. [Cited on page 22]

[54] F. Qu, Z. Wu, F. Wang, and W. Cho. A security and privacy review of vanets. *IEEE Transactions on Intelligent Transportation Systems*, 2015. [Cited on page 22]

[55] R. Song, L. Korba, and G. Yee. Pseudonym Technology for E-Service. 2006. [Cited on page 22]

[56] G. Karagiannis, O. Altintas, E. Ekici, G. J. Heijenk, B. Jarupan, K. Lin, and T. Weil. A survey and tutorial on requirements, architectures, challenges, standards and solutions. *IEEE Commun. Surveys Tuts.*, 2011. [Cited on page 23]

[57] Alex Hern. "privacy policies of tech giants 'still not gdpr-compliant". 2018. [Cited on pages 25, 36, and 182]

[58] Jon Porter. Google fined €50 million for gdpr violation in france. *The Verge*, 2019. [Cited on page 28]

[59] A. Rashid, R. Ramdhany, M. Edwards, S. M. Kibirige, A. Babar, D.Hutchison, and R. Chitchyan. Detecting and preventing data exfiltration. *CPNI*, 2014. [Cited on page 29]

[60] Seda Guerses and Jose M. del Alamo. Privacy engineering: Shaping an emerging field of research and practice. *IEEE Security & Privacy 14: 40-46*, 2016. [Cited on page 29]

[61] J J. van Rest, D. Boonstra, M. Everts, M. van Rijn, and R. van Paassen. Designing privacy-by-design. *Annual Privacy Forum. Limmasol, Cyprus.*, 2012. [Cited on page 30]

[62] D. Solove. "i've got nothing to hide" and other misunderstandings of privacy. *San Diego Law Review*, 2007. [Cited on page 32]

[63] Ann Cavoukian et al. Privacy by design: The 7 foundational principles. implementation and mapping of fair information practices. *Information and Privacy Commissioner of Ontario, Canada*, 5, 2009. [Cited on page 40]

[64] JD Meier, Alex Mackman, Michael Dunner, Srinath Vasireddy, Ray Escamilla, and Anandha Murukan. Improving web application security: threats and countermeasures. *Microsoft Corporation*, 3, 2003. [Cited on page 44]

[65] GW Van Blarkom, John J Borking, and JG Eddy Olk. Handbook of privacy and privacy-enhancing technologies. *Privacy Incorporated Software Agent (PISA) Consortium, The Hague*, 198, 2003. [Cited on page 45]

[66] (Startpage. The world's most private search engine. *https://www.startpage.com/*, 2019. [Cited on page 46]

[67] (Duckduckgo. The search engine that doesn't track you. *https://duckduckgo.com/*, 2019. [Cited on page 46]

[68] (Disconnect. Stop search engines from tracking your searches. *https://search.disconnect.me/*, 2019. [Cited on page 46]

[69] Oscobo. Search the web. *https://www.oscobo.com/*, 2019. [Cited on page 46]

[70] M. Bezzi. An information theoretic approach for privacy metrics. 2010. [Cited on page 46]

[71] K. Chatzikokolakis, C. Palamidessi, and M. Stronati. Constructing elastic distinguishability metrics for location privacy. *Proceedings on Privacy Enhancing Technologies 2015*, 2015. [Cited on page 46]

[72] Samarati P. and Sweeney L. Generalizing data to provide anonymity when disclosing information. *PODS*, 1998. [Cited on page 46]

[73] H. Troy. Have i been pwned? *https://haveibeenpwned.com*, 2018. [Cited on page 46]

[74] M. Ashwin, J. Gehrke, D. Kifer, and M. Venkitasubramaniam. e-diversity: Privacy beyond k-anonymity. *IEEE*, 2006. [Cited on page 46]

[75] Keerthana Rajendran, Manoj Jayabalan, and Muhammad Ehsan Rana. A study of k-anonymity, l-diversity, t-closeness techniques focusing medical data. *IJCSNS International Journal of Computer Science and Network Security*, 2017. [Cited on page 47]

[76] M. Ashwin, X. He, and M. Hay. Differential privacy in the wild: A tutorial on current practices and open challenges. *Proceedings of the 2017 ACM International Conference on Management of Data. ACM*, 2017. [Cited on page 47]

[77] J. Bonneau and S. Preibusch. The privacy jungle: On the market for data protection in social networks. *Economics of information security and privacy.*, 2010. [Cited on pages 47, 87, and 106]

[78] Saraswathy Shamini Gunasekaran Thinakaran Kavenesh, Jaspaljeet Singh Dhillon and Lim Fung Chen. A conceptual privacy framework for privacy-aware iot health applications. *In 6th International Conference on Computing and Informatics (pp. 175-183)*, 2017. [Cited on page 48]

[79] Xiaoyin Wang Mitra Bokaei Hosseini James Hester Ram Krishnan Jaspreet Bhatia Travis D. Breaux Slavin, Rocky and Jianwei Niu. Toward a framework for detecting privacy policy violations in android application code. *In Proceedings of the 38th International Conference on Software Engineering, pp. 25-36.)*, 2016. [Cited on page 48]

[80] Ciaran McCormick Arosha K. Bandara Blaine A. Price Perera, Charith and Bashar Nuseibeh. Privacy-by-design framework for assessing internet of things applications and platforms. *In Proceedings of the 6th International Conference on the Internet of Things, pp. 83-92.)*, 2016. [Cited on page 48]

[81] Wendy Rowan Laura Lynch OConnor, Yvonne and Ciara Heavin. Privacy by design: informed consent and internet of things for smart health. *Procedia computer science 113: 653-658*, 2017. [Cited on page 48]

[82] Rosalie Aroni Minichiello, Victor and Victor Minichiello. In-depth interviewing: Researching people. *Longman Cheshire*, 1990. [Cited on page 50]

[83] S. A McLeod. Qualitative vs. quantitative research. simply psychology. *https://www.simplypsychology.org/qualitative-quantitative.html*, 2019. [Cited on pages 50 and 51]

[84] Ajay K. Garg Arya, Satyendra and Rakesh K. Mudgal. Social implications of carbon credit trading - a case study. *In Emerging Challenges in Business, Optimization, Technology, and Industry,pp. 145-151. Springer, Cham.*, 2018. [Cited on page 50]

[85] John Dudovskiy. The ultimate guide to writing a dissertation in business studies: A step-by-step assistance. *Pittsburgh, USA*, 2016. [Cited on page 51]

[86] FPF. Fpf mobile apps study. *https://fpf.org/wp-content/uploads/Mobile-Apps-Study-June-2012.pdf*, 2012. [Cited on page 51]

[87] A. Rashid. P. Anthonysamy, P. Greenwood. Social Networking Privacy: Understanding the Disconnect from Policy to Controls. *IEEE Computer*, 2013. [Cited on pages 62 and 77]

[88] Preserve EUProject. Preparing Secure Vehicle-to-X Communication Systems (PRESERVE) Project. *V2X Security Architecture*, 2014. [Cited on page 62]

[89] PRECIOSA. The PRivacy Enabled Capability In Co-Operative Systems and Safety Applications project (PRECIOSA) . 2007. [Cited on page 63]

[90] D. Chaum. Security without identification: Transaction systems to make big brother obsolete. *Communications of the ACM*, 1985. [Cited on page 63]

[91] J. Petit, F. Schaub, M. Feiri, and F. Kargl. Pseudonym schemes in vehicular networks: A survey. *Communications Surveys Tutorials, IEEE*, 2015. [Cited on pages 63 and 74]

[92] B. Wiedersheim, Z. Ma, F. Kargl, and P. Papadimitratos. Privacy in inter-vehicular networks: Why simple pseudonym change is not enough. *In Wireless On-demand Network Systems and Services (WONS)*, 2010. [Cited on page 63]

[93] Z. Ma, F. Kargl, and M. Weber. Measuring long-term location privacy in vehicular communication systems. *Computer Communications*, 2010. [Cited on page 63]

[94] J. Freudiger, M.H. Manshaei, J.P.Hubaux, and D.C. Parkes. On non-cooperative location privacy: a game-theoretic analysis. *ACM Conference on Computer and Communications Security.*, 2009. [Cited on page 63]

[95] Think insurance. Think:insurance. *http://www.think-ins.co.uk/young-driver-insurance.php. Retrieved May 2017*, 2017. [Cited on pages 65, 67, and 68]

[96] Marmalade insurance. Marmalade insurance. 2017. [Cited on pages 67 and 99]

[97] SmartMiles. Smartmiles insurance. 2017. [Cited on pages 67 and 68]

[98] InsureTheBox insurance. InsureTheBox insurance. 2017. [Cited on pages 67, 68, and 99]

[99] DrivePlus insurance. DrivePlus insurance. 2017. [Cited on pages 67 and 99]

[100] Pauline Anthonysamy, Matthew John Edwards, Chris Weichel, and Awais Rashid. Inferring Semantic Mapping Between Policies and Code: The Clue is in the Language. *ESSoS 2016*, 2016. [Cited on page 74]

[101] Awais Rashid. Pauline Anthonysamy, Phil Greenwood. A Method for Analysing Traceability between Privacy Policies and Privacy Controls of Online Social Network. *APF 2012*, 2012. [Cited on page 74]

[102] I-Min Lee, Eric J Shiroma, Felipe Lobelo, Pekka Puska, Steven N Blair, Peter T Katzmarzyk, Lancet Physical Activity Series Working Group, et al. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *The lancet*, 380(9838):219–229, 2012. [Cited on page 76]

[103] Borja Martínez-Pérez, Isabel De La Torre-Díez, and Miguel López-Coronado. Privacy and security in mobile health apps: a review and recommendations. *Journal of medical systems*, 39(1):181, 2015. [Cited on page 77]

[104] Xitong Guo, Xiaofei Zhang, and Yongqiang Sun. The privacy–personalization paradox in mhealth services acceptance of different age groups. *Electronic Commerce Research and Applications*, 16:55–65, 2016. [Cited on page 77]

[105] Shifali Arora, Jennifer Yttri, and Wendy Nilsen. Privacy and security in mobile health (mhealth) research. *Alcohol research: current reviews*, 36(1):143, 2014. [Cited on page 77]

[106] He Debiao. A provably-secure cross-domain handshake scheme with symptoms-matching for mobile healthcare social network. *IEEE Transactions on Dependable and Secure Computing 15.4 (2016): 633-645.*, 2016. [Cited on page 77]

[107] Yao Lin. Privacy preservation in publishing electronic health records based on perturbation. *International Conference on Security and Privacy in New Computing Environments. Springer, Cham, 2019*, 2019. [Cited on page 77]

[108] Pawan Bhaladhare and Devesh Jinwala. Novel approaches for privacy preserving data mining in k-anonymity model. *Journal of Information Science and Engineering*, 2016. [Cited on page 77]

[109] R. A. Thanseeh Deshkar Sankalp and Varun G. Menon. A review on iot based m-health systems for diabetes. *International Journal of Computer Science and Telecommunications 8.1 (2017): 13-18*, 2017. [Cited on page 78]

[110] Lattie E. G. Kording K. P. Mohr D. C. Saeb, S. Mobile phone detection of semantic location and its relationship to depression and anxiety. *JMIR mHealth and uHealth 5, no. 8 (2017): e112.*, 2017. [Cited on page 78]

[111] Dayton Ward. Fit to be spied: Fitness trackers and opsec risks. *NCO Journal, Army University Press*, 2018. [Cited on page 78]

[112] Sam Schechner. You give apps sensitive personal information. then they tell facebook. *Wall Street Journal*, 2019. [Cited on page 82]

[113] Kari Paul. Fitness and health apps may be sharing the most private details about your life. *https://www.marketwatch.com/*, 2019. [Cited on page 82]

[114] M. Finley Frazee, J. and J. J. Rohack. mhealth and unregulated data: is this farewell to patient privacy. *Ind. Health L. Rev. 13 (2016): 384*, 2016. [Cited on page 85]

[115] AppBrain Stats. Most popular google play categories, 2015. [Cited on page 89]

[116] Aviva. Telematics - young drivers. *https://broker.aviva.co.uk/products/marketplace/telematics-youngdrivers/*, 2019. [Cited on page 99]

[117] Tomtom. Tomtom curfer. *https://www.tomtom.com/en_gb/sat-nav/curfer/*, 2019. [Cited on page 99]

[118] Hastings. Hastings direct smart miles - black box (telematics) car insurance. *https://www.hastingsdirect.com/car-insurance/smartmiles.shtml*, 2019. [Cited on page 99]

[119] Flo. Driving you to a better driver. *https://www.decosdrivinginsights.com/?lang=nl*, 2019. [Cited on page 99]

[120] Smart Wheels. Smart wheels - young driver insurance. *https://www.morethan.com/car-insurance/young-driver-insurance/*, 2019. [Cited on page 99]

[121] Seat. Seat telematics. *https://www.insurewithseat.co.uk/*, 2019. [Cited on page 99]

[122] VW. Volkswagen telematics. *https://www.insurewithvolkswagen.co.uk/*, 2019. [Cited on page 99]

[123] (GPEN). Results of the 2016 global privacy enforcement network sweep. *https://www.priv.gc.ca/en/opc-news/news-and-announcements/2016/bg_160922/*, 2019. [Cited on page 103]

[124] (GDPR). Art. 45 gdpr transfers on the basis of an adequacy decision. *https://gdpr-info.eu/art-45-gdpr/*, 2019. [Cited on page 105]

[125] (GDPR). Art. 46 gdpr transfers subject to appropriate safeguards. *https://gdpr-info.eu/art-46-gdpr/*, 2019. [Cited on page 105]

[126] Bin Liu, Deguang Kong, Lei Cen, Neil Zhenqiang Gong, Hongxia Jin, and Hui Xiong. Personalized mobile app recommendation: Reconciling app functionality and user privacy preference. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, pages 315–324. ACM, 2015. [Cited on page 113]

[127] Sunil Patil, Dimitris Potoglou, Hui Lu, Neil Robinson, and Peter Burge. Trade-off across privacy, security and surveillance in the case of metro travel in europe. *Transportation Research Procedia*, 1(1):121–132, 2014. [Cited on page 113]

[128] Xi He, Ashwin Machanavajjhala, and Bolin Ding. Blowfish privacy: Tuning privacy-utility trade-offs using policies. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, pages 1447–1458. ACM, 2014. [Cited on page 113]

[129] William Enck, Peter Gilbert, Seungyeop Han, Vasant Tendulkar, Byung-Gon Chun, Landon P Cox, Jaeyeon Jung, Patrick McDaniel, and Anmol N Sheth. Taintdroid: an information-flow tracking system for realtime privacy monitoring on smartphones. *ACM Transactions on Computer Systems (TOCS)*, 32(2):5, 2014. [Cited on page 113]

[130] Stéphanie Lefevre, Jonathan Petit, Ruzena Bajcsy, Christian Laugier, and Frank Kargl. Impact of v2x privacy strategies on intersection collision avoidance systems. In *Vehicular Networking Conference (VNC), 2013 IEEE*, pages 71–78. IEEE, 2013. [Cited on page 115]

[131] Jonathan Petit and Steven E Shladover. Potential cyberattacks on automated vehicles. *IEEE Trans. Intelligent Transportation Systems*, 16(2):546–556, 2015. [Cited on page 115]

[132] Sebastian Derikx, Mark de Reuver, and Maarten Kroesen. Can privacy concerns for insurance of connected cars be compensated? *Electronic Markets*, 26(1):73–81, 2016. [Cited on page 116]

[133] G. Danezisn, J. Domingo-Ferrer, M. Hansen, J. H. Hoepman, D. L. Metayer, R. Tirtea, and S. Schiffner. Privacy and Data Protection by Design-from policy to engineering. *enisa.europa.eu*, 2015. [Cited on page 117]

[134] K. Wuyts, R. Scandariato, W Joosen, M. Deng, and B. Preneel. LINDDUN Privacy Threat Modeling. *LINDDUN*, 2018. [Cited on page 118]

[135] LINDDUN privacy threat modeling: a tutorial. *Leuven*, 2015. [Cited on pages 118, 120, and 125]

[136] Challenges of privacy requirements modelling in V2X applications: A telematic insurance case study. *IEEE 25th International Requirements Engineering Conference Workshops (REW)*, 2017. [Cited on pages 118 and 125]

[137] Ibm rational doors: A requirements management solution that helps you capture, trace, analyze and manage systems and advanced it application development. 2018. [Cited on page 119]

[138] www.irise.com. 2018. [Cited on page 119]

[139] T. Coventry. Requirements management - planning for success!: techniques to get it right when planning requirements. *PMI Global Congress 2015 - EMEA; https://www.pmi.org/learning/library/requirements-management-planning-for-success-9669*, 2019. [Cited on page 119]

[140] S Dasanayake, S Aaramaa, J Markkula, and M Oivo. Impact of requirements volatility on software architecture: How do software teams keep up with ever-changing requirements? *Journal of Software: Evolution and Process*, 2019. [Cited on page 119]

[141] K. Crawford and J. Schultz. Big data and due process: Toward a framework to redress predictive privacy harms. *Boston College Law Review,*, 2014. [Cited on page 120]

[142] State of California. California legislative information, 2018. [Cited on page 129]

[143] mysugr. . *https://mysugr.com/us-contest-show-us-sugr*, 2018. [Cited on page 130]

[144] OASIS Privacy Management Reference Model (PMRM). 2018. [Cited on page 138]

[145] B. Terhorst and Arnaud van den Berg. Why mobile app architecture is vital for mobile app development?, 2018. [Cited on page 141]

[146] ISO. The international organization for standardization, usability definitions and concepts. *ISO 9241-11 Ergonomics of Human-system Interaction*, 2018. [Cited on page 162]

[147] P. W Jordan. An introduction to usability. *Taylor and Francis, London*, 1998. [Cited on page 162]

[148] Usability.org. Usability testing. *U.S Department of Health & Human Services*, 2019. [Cited on page 163]

[149] John Brooke. Sus - a quick and dirty usability scale: Smart phone applications for people with brain injury. *www.TBIStaffTraining.info*, 2019. [Cited on pages 169 and 172]

[150] UIUX Trend. Measuring and interpreting system usability scale (sus), https://uiuxtrend.com/measuring-system-usability-scale-sus/. *UX Research 2019*, 2019. [Cited on pages 169 and 170]

[151] Kortum P. & Miller J. Bangor, A. Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies 4, no. 3 : 114-123.*, 2009. [Cited on page 170]

[152] William G. Marchal Lind, Douglas A. and Samuel Adam Wathen. Statistical techniques in business & economics. *McGraw-Hill/Irwin*, 2015. [Cited on page 170]

[153] Wolfgang Stoettner. Statistics formulae collection; collection of statistics formulae taken from the perennial text book lind, douglas a. et. al. (2015): Statistical techniques in business and economics, 16 ed. *Overleaf.com*, 2019. [Cited on page 170]

[154] Statistics Solutions. Mann-whitney u test. *statisticssolutions.com*, 2019. [Cited on page 171]

[155] Social science statistics. *https://www.socscistatistics.com/tests/mannwhitney/*, 2019. [Cited on page 171]

[156] Finstad Kraig. The usability metric for user experience. *Interacting with Computers 22.5: 323-327*, 2010. [Cited on page 172]

[157] James R. Lewis, Brian S. Utesch, and Deborah E. Maher. Umux-lite: when there's no time for the sus. *ACM, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.*, 2013. [Cited on page 172]

[158] Sauro Jeff. Supr-q: a comprehensive measure of the quality of the website user experience. *Journal of usability studies 10.2: 68-86.*, 2015. [Cited on page 172]

[159] Net promoter score. *http://www.netpromoter.com*, 2019. [Cited on page 172]

[160] Ari Ezra Waldman. Privacy, sharing, and trust: The facebook study. *Case W. Res. L. Rev.*, 67:193, 2016. [Cited on page 186]

[161] Privacy policy template generator. *https://www.privacypolicytemplate.net/*, 2019. [Cited on page 187]

[162] Avantika Monnappa. How facebook is using big data - the good, the bad, and the ugly. *https://www.simplilearn.com/how-facebook-is-using-big-data-article*, 2019. [Cited on page 187]

[163] Compliance Junction. Beuc: Large tech companies privacy policies are not gdpr compliant. *https://www.compliancejunction.com/beuc-large-tech-companies-privacy-policies-are-not-gdpr-compliant/*, 2018. [Cited on page 188]

Appendix

App Permissions

Table 1: App permissions

Permissions	7 cups	Depression CBT	Operation Reach Out	Mood Scanner	Fight Depression naturally	Diabetes PA	Dexcom	Diabetes Tracker	MySugr	Glooko	Fitbit	Nike training club	Freelotics Bodyweight	Sworkit	Garmin Connect	Instant Heart rate	Pulse Point	Nokia Health Mate	Daily Cardio Workout	Cardiogram	Idealweight	Losoit	Weight Watchers	Noom coach	Calorie Counter	Medscape	PEPID	UpToDate	Amnon Doximity	HealthyQxMD	Number of Permissions	Number of Apps	Percent
Read your own contact card						✓						✓				✓																3	30	10%
Add or remove accounts												✓				✓													✓			3	30	10%
Read sensitive log data																															1	30	3%
Find accounts on the device				✓		✓			✓			✓		✓		✓	✓	✓					✓		✓		✓				11	30	37%
Read your own contact card						✓						✓				✓															4	30	13%
Read your contacts											✓																				5	30	17%
Modify your contacts			✓																												1	30	3%
Read phone status and identity	✓	✓							✓	✓	✓	✓			✓	✓	✓	✓		✓			✓	✓	✓				✓		18	30	60%
Access USB storage file system																															0	30	0%
Directly call phone numbers				✓				✓							✓																3	30	10%
Read call log															✓																1	30	3%
Write call log															✓																1	30	3%
Photos/Media/Files																						✓									1	30	3%
Access USB storage file system															✓								✓								2	30	7%
Read the contents of your USB storage	✓		✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	✓	✓		✓	✓	✓	✓	26	30	87%
Modify or delete the contents of your USB storage	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	✓	✓		✓	✓	✓	✓	26	30	87%
Read the contents of your USB storage	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	✓	✓		✓	✓	✓	✓	25	30	83%
Modify or delete the contents of your USB storage	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	✓	✓		✓	✓	✓	✓	25	30	83%
Read phone status and identity	✓	✓				✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓		✓	✓	✓	✓	17	30	57%
Approximate location (network-based)		✓							✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓						14	30	47%
Precise location (GPS and network-based)									✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓						14	30	47%
Access extra location provider commands																															2	30	7%
Receive text messages(sms)									✓						✓		✓							✓							3	30	10%
Read your text messages(SMS or MMS)															✓																1	30	3%
Receive text messages(mms)															✓																1	30	3%
Send sms messages																															0	30	0%
Read Calendar events plus confidential information																															0	30	0%
Take pictures and videos												✓				✓		✓		✓		✓	✓	✓							11	30	37%
Record audio										✓													✓								3	30	10%
View WiFi connections												✓	✓		✓			✓		✓			✓		✓			✓	✓	✓	11	30	37%
Heart rate monitors																✓															2	30	7%
Read Calendar events plus confidential information															✓																2	30	7%
Receive data from internet	✓		✓	✓	✓	✓			✓				✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	24	30	80%
View network connections	✓	✓		✓	✓	✓			✓				✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	24	30	80%
Pair with Bluetooth devices															✓																8	30	27%
Access Bluetooth settings									✓	✓	✓				✓					✓			✓	✓	✓						7	30	23%
Control Near-Field Communication											✓																				1	30	3%
Read Google service configuration											✓				✓			✓													5	30	17%
Toggle sync on and off												✓				✓							✓								2	30	7%
Download files without notification												✓				✓							✓								4	30	13%
Read sync statistics and settings												✓											✓								2	30	7%
Full network access	✓		✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	28	30	93%
Use accounts on the device					✓									✓																	6	30	20%
Control flashlight																✓				✓					✓						5	30	17%
Control media playback and metadata access															✓																1	30	3%
Smart card services permissions label															✓																1	30	3%
Send sticky broadcast								✓							✓														✓		3	30	10%
Change network connectivity															✓										✓						2	30	7%
Connect and disconnect form WiFi															✓										✓						2	30	7%
Change your audio settings															✓																1	30	3%
Run at startup									✓				✓	✓	✓		✓	✓					✓		✓						10	30	33%
Draw over other apps										✓					✓																2	30	7%
Control vibration													✓	✓	✓		✓	✓					✓	✓	✓				✓		11	30	37%
Prevent device from sleeping								✓		✓			✓	✓	✓		✓	✓	✓				✓	✓					✓		12	30	40%
Modify system settings															✓																2	30	7%
Google play licence check													✓	✓	✓																2	30	7%
Reorder running apps													✓																		1	30	3%
Prevent device from sleeping			✓	✓			✓										✓												✓	✓	7	30	23%
Close other apps								✓																							1	30	3%
Install shortcuts								✓																							3	30	10%
Manage document storage																													✓		1	30	3%
Unistall shortcuts																															1	30	3%
Create accounts and set passwords																													✓		1	30	3%

eHealth Class diagram

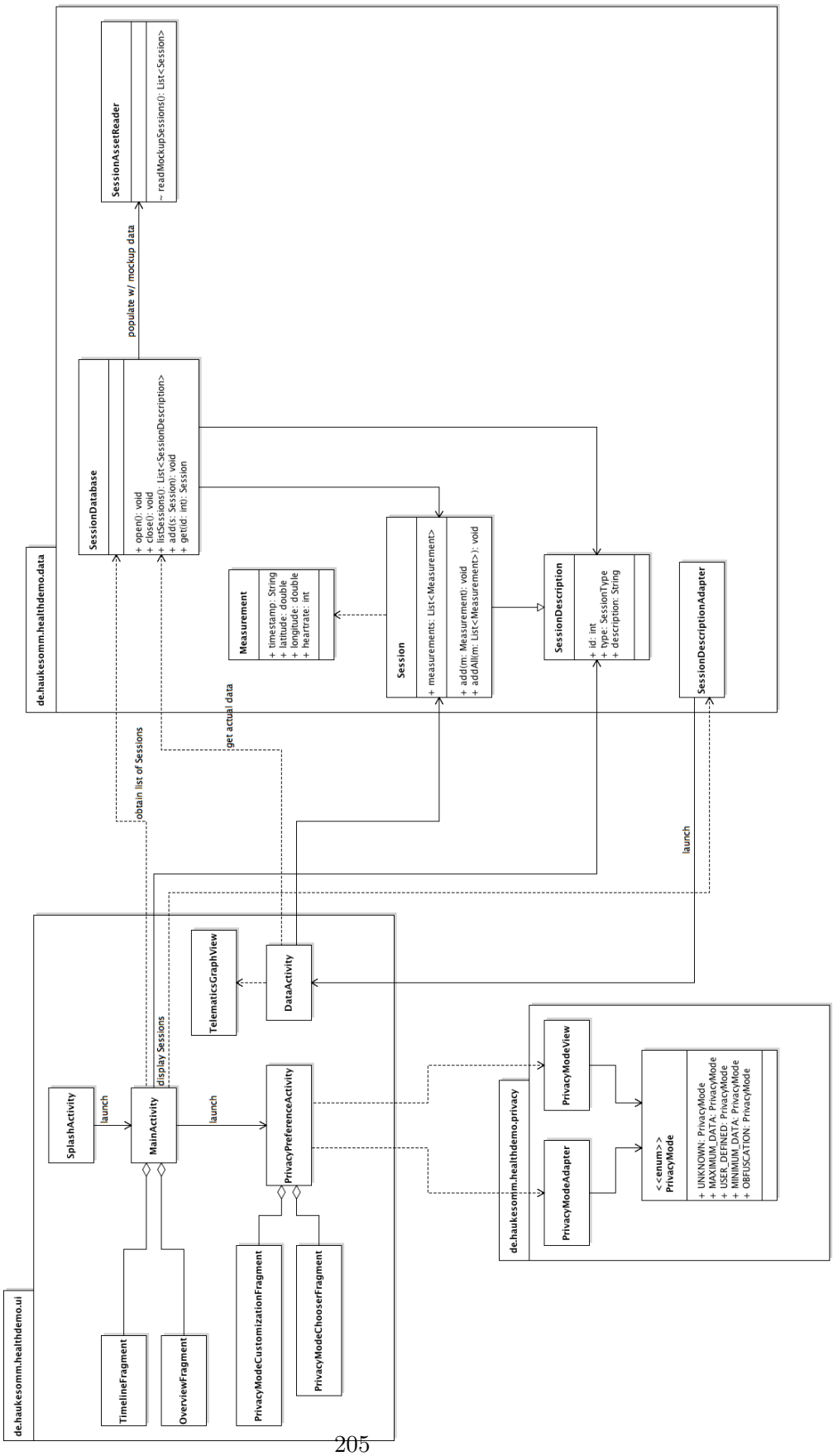


Figure 1: eHealth Class diagram

System Usability Scores

Table 2: System Usability Scores of the Ehealth app

ID	ID 1	ID 2	ID 3	ID 4	ID 5	ID 6	ID 7	ID 8	ID 9	ID 10	ID 11	ID 12	ID 13	ID 14	ID 15	ID 16	ID 17	ID 18	ID 19	ID 20	ID 21	ID 22	ID 23	ID 24	ID 25	ID 26	ID 27	ID 28	ID 29	ID 30	ID 31	ID 32	ID 33	ID 34	ID 35	ID 36	ID 37	ID 38	ID 39	ID 40	ID 41	ID 42	ID 43	ID 44	ID 45	ID 46	ID 47	ID 48	ID 49	ID 50
I think that I would like to use this app frequently.	4	4	5	4	5	4	4	4	4	5	4	4	4	4	5	5	2	5	5	5	5	5	4	4	5	4	4	5	5	3	5	4	4	5	5	4	4	4	5	4	4	5	5	4	4	5	5	4	4	4
I thought that there was too much inconsistency in this app.	3	2	1	2	2	2	2	3	3	3	3	2	2	4	3	4	2	3	2	2	3	4	3	2	3	1	2	1	2	3	3	4	3	4	1	2	3	3	2	4	4	3	3	2	3	4	2	3	1	2
I thought the app was easy to use.	5	4	5	4	5	4	5	3	4	3	4	4	5	4	3	5	3	4	4	4	5	4	5	3	4	5	3	5	4	3	4	4	5	5	5	3	4	5	4	4	4	5	5	4	5	5	4	4	5	4
I think I would need the support of a technical person to use.	2	3	5	4	4	4	4	4	3	3	4	3	4	2	3	4	4	3	4	3	2	4	2	3	3	4	3	2	1	4	3	4	4	5	3	2	3	4	4	3	5	5	5	4	2	3	1	2	3	2
I found the various functions in this app were well integrated.	4	5	5	5	4	4	4	4	3	3	4	4	3	4	4	3	4	4	5	5	4	4	3	5	5	4	3	4	5	5	5	5	4	3	4	4	5	4	3	5	5	5	5	5	4	3	5	3	4	5
I thought there was a lot of consistency in using this app.	1	2	2	2	4	3	3	2	1	3	4	4	4	4	3	3	1	4	3	4	4	4	4	4	4	3	4	5	2	3	4	3	5	4	4	3	2	5	3	4	3	3	3	3	3	2	2	2	3	1
I needed to learn a lot of things before I could get going with this app.	1	2	2	2	4	2	3	3	2	1	3	4	4	2	5	3	3	1	4	3	4	4	4	4	5	5	4	4	2	2	1	2	3	4	5	3	2	5	3	4	3	5	5	3	3	2	2	2	3	1
I found the various functions of the app were well integrated.	4	3	4	4	4	4	4	4	4	3	4	4	3	2	4	3	4	4	4	5	3	5	5	4	5	4	5	4	1	5	3	4	4	2	4	5	4	4	4	5	4	4	4	4	4	5	4	5	3	4
I think that I would like to use this app frequently.	4	4	5	4	3	4	5	4	4	4	3	4	4	4	3	5	4	3	3	4	5	4	4	4	4	4	4	5	4	4	5	5	4	4	5	4	3	5	4	5	4	3	5	4	5	4	5	4	5	4
I would imagine that most people would learn to use this system very quickly.	4	4	5	4	3	4	5	4	4	4	3	4	4	4	3	5	4	3	3	4	5	4	4	4	4	4	4	5	4	4	5	5	4	4	5	4	3	5	4	5	4	3	5	4	5	4	5	4	5	4
I think that I would need the support of a technical person to be able to use this app.	1	2	2	2	3	2	2	2	1	3	3	3	2	3	3	3	2	3	2	2	3	2	1	3	2	3	2	1	1	4	2	3	3	1	3	2	3	2	2	1	2	3	2	2	2	1	3	1	2	2
SUS SCORE	82.5	72.5	85	72.5	67.5	70	67.5	67.5	80	50	55	57.5	72.5	57.5	52.5	62.5	75	57.5	72.5	65	67.5	62.5	70	60	70	75	70	92.5	67.5	50	62.5	67.5	60	65	82.5	77.5	62.5	65	72.5	62.5	57.5	62.5	80	70	75	70	90	77.5	80	77.5
SUS TOTAL	68.9
AVERAGE

Security Lancaster | Lancaster University

Consent Form

Study Title: Privacy preservation in mobile apps – Usability study.

The purpose of this consent form is to check that you are aware of your rights, understand what will be required of you and agree to take part in the study. If you have any concerns before signing the consent form please contact the principal investigator, [Sarah Kibirige Mukisa].

	Please tick each Box.
1. I confirm that I have read the information sheet and fully understand what is expected of me within this study	☐
2. I confirm that I have had the opportunity to ask any questions about the research and have them answered satisfactorily.	☐
3. I understand that my participation is voluntary and that I am free to withdraw anytime within two (2) weeks after the interview without giving any reason.	☐
4. I understand that the information collected during the study will be pooled with that of other participants, anonymised and aggregated before being published.	☐
5. I understand that once my data have been anonymised and incorporated into themes it might not be possible for it to be withdrawn, though every attempt will be made to extract my data, up to the point of publication.	☐
6. I am satisfied that the information I provide will be treated confidentially by the researchers.	☐
7. I agree that quotations from the interviews can be used in the project reports and in other publications (if applicable). I understand that my quotations will be used anonymously.	☐
8. By ticking this box, you consent to taking part in the current study.	☐

_____	_____	_____
Name of Participant	Date	Signature

Figure 2: Participant consent form

For further information about how Lancaster University processes personal data for research purposes and your data rights please visit our webpage:
www.lancaster.ac.uk/research/data-protection

Participant Information Sheet

I am a PhD researcher at Lancaster University and I would like to invite you to take part in a research study about privacy preservation in future mobile applications.

Please take time to read the following information carefully before you decide whether you wish to take part.

What is the study about?

This study is aimed at designing better privacy aware mobile applications. There has been a surge in privacy breaches in mobile applications ranging from data exposure to data manipulation of the data collected by mobile applications as in the case of Facebook and Yahoo data leaks. Previous research in this area has looked at data breaches and found that users are concerned about the data collected. However, at present there are no solutions proposed to tackle this gap. Taking part in this research aids researchers in designing better privacy aware mobile apps.

We have designed two demo mock-up apps to investigate privacy infringement in current mobile apps. We designed a mock-up Telematics insurance app and an eHealth fitness app. Telematics insurance is a type of car insurance which uses telematics technology to determine the cost of the insurance policy used. Instead of using a yearly or monthly policy that is based on risk and factors like, where you live, make of the car, model of the car, telematics insurance is different. It is based on the driving behaviour using a black box that monitors a user's driving style.

The eHealth app is used to determine how a user keeps fit by making fitness activities like riding a bicycle and monitoring the heart rate. Fitness apps are used to encourage users to improve on their fitness activities.

If you take part in this study, you will be asked to investigate and give us feedback on the usability of the Telematics and eHealth fitness apps. The apps do not collect any personal data of any activities you will perform.

Why have I been invited?

I have approached you because I want to get your ideas and views on how to improve mobile applications on the market today. I want to know what you consider important when using the mobile applications to be tested, the features you would like to see in mobile apps with regards to data collection and privacy preservation. In short, I want you to help me design a better interface to solve users' privacy in mobile applications. This is an open invitation to people who use mobile apps on a regular basis and have submitted private data to mobile applications.

The demo apps have inbuilt privacy preserving features which you will investigate and give feedback subsequently. The apps are on Github a software source code management service. You will be given the links to the apps which you will download on your android based phone.

I will be very grateful if you agree to take part in this study.

Figure 3: Participant Information sheet