Secure and Privacy-aware Data Collection and Processing in Mobile Health Systems

Leonardo Horn Iwaya

Faculty of Health, Science and Technology

Computer Science

LICENTIATE THESIS | Karlstad University Studies | 2016:47

Secure and Privacy-aware Data Collection and Processing in Mobile Health Systems

Leonardo Horn Iwaya

LICENTIATE THESIS | Karlstad University Studies | 2016:47

Secure and Privacy-aware Data Collection and Processing in Mobile Health Systems

Leonardo Horn Iwaya

LICENTIATE THESIS

Karlstad University Studies | 2016:47

urn:nbn:se:kau:diva-46982

ISSN 1403-8099

ISBN 978-91-7063-730-8

Distribution:
Karlstad University
Faculty of Health, Science and Technology
Department of Mathematics and Computer Science
SE-651 88 Karlstad, Sweden
+46 54 700 10 00

Print: Universitetstryckeriet, Karlstad 2016

WWW.KAU.SE

Secure and Privacy-aware Data Collection and Processing in Mobile Health Systems

Leonardo Horn Iwaya
Department of Mathematics and Computer Science

Abstract

Healthcare systems have assimilated information and communication technologies in order to improve the quality of healthcare and patient's experience at reduced costs. The increasing digitalization of people's health information raises, however, new threats regarding information security and privacy. Accidental or deliberate data breaches of health data may lead to societal pressures, embarrassment and discrimination. Information security and privacy are paramount to achieve high quality healthcare services, and further, to not harm individuals when providing care. With that in mind, we give special attention to the category of Mobile Health (mHealth) systems. That is, the use of mobile devices (e.g., mobile phones, sensors, PDAs) to support medical and public health. Such systems, have been particularly successful in developing countries, taking advantage of their flourishing mobile market and their need to expand the coverage of primary healthcare programs. Many mHealth initiatives, however, fail to address security and privacy issues. This, coupled with the lack of specific legislation for privacy and data protection in these countries, increases the risk of harm to individuals. The overall objective of this thesis is to enhance knowledge regarding the design of security and privacy technologies for mHealth systems. In particular, we deal with mHealth Data Collection Systems (MDCSs), which consists of mobile devices for collecting and reporting health-related data, replacing paper-based approaches for health surveys and surveillance. This thesis consists of publications contributing to mHealth security and privacy in various ways: with a comprehensive literature review about mHealth in Brazil; with the design of a security framework for MDCSs (SecourHealth); with the design of a MDCS (GeoHealth); with the design of Privacy Impact Assessment template for MDCSs; and with the study of ontology-based obfuscation and anonymisation functions for health data.

Keywords: Mobile health, information security, data privacy, data collection, personal health data.

Acknowledgements

This licentiate thesis would never be accomplished without the help of other people.

First of all, I would like to thank my family for all the encouragement towards pursuing an academic life. Education has always been a priority, seen as a platform for achieving freedom and fulfillment. I am extremely thankful for all the support invested in me in order to acquire always the best education opportunities throughout my life.

I am also grateful to have Simone Fischer-Hübner, Leonardo Martucci, and Rose-Mharie Åhlfeldt as my supervisor and co-supervisors. Your support has been essential to help me master science at its highest level. Furthermore, your teachings leave an indelible mark, that positively influence my career as a researcher (and future educator).

I would like to thank academics that collaborated with this research, during the period of my master studies, at the University of São Paulo (USP), as well as during my doctoral studies at Karlstad University (KAU). Also, I thank my colleagues from the department of Computer Science at KAU, for the many fruitful discussions and feedback during *fika* and for creating such a wonderful working environment. I also extend my thanks to my dear friends that from near and far helped me to keep chasing my dreams.

This research has been partly funded by: the European Commission under the Seventh Framework Programme in the project SmartSociety; the Swedish Knowledge Foundation in project High Quality Networked Services in a Mobile World (HITS); and, the Nordic Digital Health Center (NDHC), a cooperative project between the county of Värmland, the municipality of Karlstad, Karlstad University and Stiftelsen Compare Karlstad.

Karlstad (Sweden), November 2016 Leonardo Horn Iwaya

List of Appended Papers

I. L.H. Iwaya, M.A.L. Gomes, M.A. Simplício, T.C.M.B. Carvalho, C.K. Dominicini, R.R.M. Sakuragui, M.S. Rebelo, M.A. Gutierrez, M. Näslund and P. Håkansson. Mobile Health in Emerging Countries: A Survey of Research Initiatives in Brazil. International Journal of Medical Informatics, Volume 82, Issue 5, May 2013, Pages 283-298, ISSN 1386-5056.

II. M.A. Simplício, L.H. Iwaya, B.M. Barros, T.C.M.B. Carvalho and M. Näslund. SecourHealth: A Delay-Tolerant Security Framework for Mobile Health Data Collection. IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 2, pp. 761-772, March 2015.

III. J.H.G. Sá, M.S. Rebelo, A. Brentani, S. Grisi, L.H. Iwaya, M.A. Simplicio Jr., T.C.M.B. Carvalho and M.A. Gutierrez. Georeferenced and Secure Mobile Health System for Large Scale Data Collection in Primary Care. International Journal of Medical Informatics, Volume 94, p. 91-99, 2016.

IV. L.H. Iwaya, L.A. Martucci and S. Fischer-Hübner. Towards a Privacy Impact Assessment Template for Mobile Health Data Collection Systems. Proceedings of the 5th International Conference on M4D Mobile Communication Technology for Development: M4D 2016, General Tracks.

V. L.H. Iwaya, F. Giunchiglia, L.A. Martucci, A. Hume, S. Fischer-Hübner and R. Chenu-Abente. Ontology-based Obfuscation and Anonymisation for Privacy: A Case Study on Healthcare. Privacy and Identity Management. Time for a Revolution? 10th IFIP WG 9.2, 9.5, 9.6/11.7, 11.4, 11.6/SIG 9.2.2 International Summer School, Edinburgh, UK, August 16-21, 2015, p. 343–358, Revised Selected Papers.

Comments on my Participation

Paper I The survey was proposed by me. I participated in the entire writing process but specially Sections *"Mobile health in Brazil"* and *"Analysis and Discussion"*. I am also the main author of the publication.

Paper II I participated in the entire writing process but specially in Section *"Implementation"*, i.e., SecourHealth's implementation and testing.

Paper III I participated in the entire writing process but specially in Subsection *"GeoHealth's security features"*.

Paper IV The PIA template was proposed by me. I participated in the entire writing process but specially in Sections *"Mobile Health Data Collection Systems"*, *"MDCS worldwide"*, *"Information Security and Privacy Perspectives"*, and *"Narrowing down"*. I am also the main author of the publication.

Paper V I participated in the entire writing process but specially in Sections *"Background and related work"* and *"Obfuscation and anonymisation for HIS"*. I am also the main author of the publication.

Other publications

- B.M. Barros, L.H. Iwaya, M.A. Simplício Jr., T.C. M. B. Carvalho, A. Méhes and M. Näslund. Classifying security threats in cloud networking. In Proceedings of the 5th International Conference on Cloud Computing and Services Science, pages 214–220, 2015.
- M.A. Simplício, T.C.M.B. Carvalho, C.K. Dominicini, P. Håkansson, L.H. Iwaya and M. Näslund. Method and apparatus for securing a connection in a communications network. Patent US 2015/0281958, October 2015.
- M. Näslund, T.C.M.B. Carvalho, L.H. Iwaya and M.A. Simplício. Encrypting and storing data, Patent US 2016/0156464, June 2016.
- L.H. Iwaya, A. Voronkov, L.A. Martucci, S. Lindskog and S. Fischer-Hübner. Firewall Usability and Visualization: A Systematic Literature Review, Karlstad University Studies, ISBN 978-91-7063-718-6, 2016.

Contents

Introductory Summary

"Freedom is the alone unoriginated birthright of man, and belongs to him by force of his humanity;"

— *The Metaphysics of Ethics. (1886)*
Immanuel Kant

1 Introduction

Healthcare has changed over the years with the incorporation information and communication technologies. This transformation also affected other aspects of healthcare, such as: the relationship between patients and doctors; the ways to deliver healthcare; and, the capacity of data analysis for clinical and research purposes. Health Informatics (HI) is now a well established research area, with many branches and ramifications. Electronic Health Records (EHRs) began to appear in early 1990s. Nowadays, EHRs are part of day-to-day reality in most health facilities around the world. Advances in telecommunications led to healthcare over distances. Telemedicine connects medical professionals in different sites, allowing tele-consultation and delivery of specialized healthcare to remote locations. With the introduction of mobile computing and wireless communication technologies, applications are now designed to run in smart-phones and to use sensor networks to monitor patients in real-time. Healthcare professionals can access all the relevant or needed data through many computer interfaces (e.g., desktops, smartphones, tablets). Likewise, the patients can have readily access to their medical journals by Internet. Essentially, information and communications technology help to improve quality of healthcare and patient's experience at reduced costs.

Notwithstanding, the security and privacy risks grew proportionally. Huge amounts of data have to be securely transmitted, processed, and stored. The disruption of communication channels can prevent patients from receiving healthcare in an emergency situation. Data breaches on patient's medical records can cause societal pressure, embarrassment and discrimination. Systems can be potentially misused in patient's detriment. Privacy infringements can be caused by, e.g., purpose misuse, vague purpose specification, lack of patient's consent, and privacy policies. And furthermore, in most countries there is no legal framework that regulates privacy and protection of personal data.

In this licentiate thesis we look at the area of Health Informatics (HI) from the standpoints of information security and privacy. More specifically, a considerable amount of this work is dedicated to mobile health (mHealth) systems. This category of applications stems from advances in mobile computing, wireless communication and global positioning systems (GPS). mHealth capitalizes on mobile phone's core utility of voice and short messaging service (SMS), as well as more complex functionalities, such as Bluetooth and mobile broadband (e.g., UMTS, GSM, LTE) [41]. mHealth works as a big umbrella term with various subcategories of applications. Solutions are developed either to support health workers on their activities or to give to individuals more control of their own health. In both cases, mHealth applications normally extend the capacity of managing (i.e., store, retrieve, transmit, manipulate, and reasoning) health data.

Several mHealth projects and initiatives exist worldwide. Thousands of mHealth applications can be found in different digital distribution platforms (e.g., Google Play and App Store). In the majority of cases, wellness and fit-

ness applications that help users to better plan their diet and exercise routines. Nevertheless, other mHealth applications can have clinical value, targeting different groups of users, such as medical professionals, elderly people, chronic patients, pregnant women, and so on. The mHealth market thrives in developed and developing countries, with different business models and contexts of use.

However, despite its potential for effectively improving quality of healthcare, many mHealth proposals do not employ robust-enough security and privacy-preserving solutions. In essence, such situation prevents real deployment of initiatives, specially in low- and middle- income countries. Thus, this is forcing promising ideas to stay in small-scale or prototypical levels, exactly where they are most needed to improve healthcare coverage, readiness and efficiency.

In this thesis we present a series of papers on new healthcare technologies. Information security and privacy received special attention in all publications. Although each of the papers has a specific aim, the thesis as a whole contributes to the state of the art on security and privacy during the collection and processing of data in mHealth systems. And it does so, by showing how to safeguard health information through the design and implementation of security and privacy-preserving mechanisms.

The remainder of this Introductory Summary is structured as follows. Section 2 provides the research background, by defining terms and concepts and technologies used in this thesis. Section 3 introduces the thesis research question. Section 4 describes the research method that was predominantly adopted in this work. Section 5 states the scientific contributions achieved with this research. Section 6 discusses the relevant related work. Section 7 contains a list of the appended publications. Section 8 presents the research conclusions and directions for future work.

2 Background

This section presents the background, terms and concepts, used in this thesis. First, some definitions for the concepts of information, privacy and security are presented. Then, the relevant types of Health Informatics (HI) for the thesis are described, namely: Electronic Health Records (EHRs), Personal Health Records (PHRs) and mobile healthcare (mHealth). At last, general issues regarding security and privacy in HI are discussed, with emphasis in mHealth Data Collection System (MDCS). It is worthy mentioning, however, that this background is not meant to be exhaustive, neither to go into the details of each concept, but rather to provide the sufficient definitions for the readers.

2.1 Information, Privacy and Security

When someone asks you a question, you may decide to respond to it or not. In any case, the answer of the question is considered *information*, i.e., knowledge

about someone or something. Besides, a few other questions could come to the respondent's mind – hopefully before he/she answers anything. Why are they asking me this question? Why they need this information? What are they going to do with it? How are they going to keep or handle it? These questions reflect the logical concern about the consequences of giving away information. In special, the consequences to someone's privacy.

The concept of privacy has not been fully consolidated in the literature. Though, a possible definition, that will be used in the scope of this thesis, was defined by Westin (1967):

> "Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others. Viewed in terms of the relation of the individual to social participation, privacy is the voluntary and temporary withdrawal of a person from the general society through physical or psychological means, either in a state of solitude or small-group intimacy or, when among larger groups, in a condition of anonymity or reserve." [55].

Privacy refers to the ability of an individual or group to seclude themselves (*physical privacy*), or the information about themselves (*information privacy*), and thereby express themselves selectively. Broadly speaking, that means that privacy has at least two categories: (i) physical privacy, and (ii) information privacy. In this thesis we focus on the latter, information privacy (also commonly referred as data privacy).

Security of information systems is another fundamental aspect of this thesis. *Computer security* is the discipline of protection of information systems (i.e., hardware, software and information) from theft, damage, disruption or misdirection. Computer security rests on three key concepts [7]:

- **Confidentiality**, the concealment of information or resources.
- **Integrity**, the trustworthiness of data resources; prevention of improper or unauthorized change.
- **Availability**, the ability to use the information or resource desired.

The concept of *information security* overlaps with *information privacy*, specially with the aspect of confidentiality. That is, the authorized access or disclosure of information, which comprises notions of secrecy, access-control, sharing, protection of information. The concept of privacy is socially constructed; privacy rights are perceived differently in countries and cultures, and systematized differently in the law. Therefore, when it comes to individuals' information, information security is one of the means to achieve information privacy. For example, individuals expect that their emails should not be read by others than the recipients, that is a privacy claim. Encryption is the security technology used as a mean to achieve such privacy goal.

For the sake of brevity, in this thesis, the terms "information security" and "information privacy" are often called simply by security and privacy. We also

avoid the use of acronyms "IS" and "IP" since they can be confused with other terms (e.g., Information Systems and Internet Protocol).

2.2 Health Informatics

Health Informatics (HI) is *"the interdisciplinary study of the design, development, adoption and application of IT-based innovations in healthcare services delivery, management and planning"* [29]. In this thesis, we mainly deal with HI applied to public health (i.e., Public Health Informatics (PHI)) and clinical medicine (i.e., Medical Informatics (MI)). Nonetheless, HI solutions can spread through a broad range of other fields, e.g., nursing, dentistry, pharmacy, biomedical research.

Lately, the term eHealth (electronic process in health) has been increasingly used to refer to health informatics using the Internet and related technologies [42, 40]. In this thesis, we are particularly interested in the following eHealth sub-categories: Electronic Health Records (EHR), Personal Health Records (PHR), and mobile health (mHealth). Each of them is defined and briefly discussed below.

2.2.1 Electronic Health Record (EHR)

EHR is probably the most widespreaded eHealth technology. According to the Healthcare Information and Management Systems Society (HIMSS):

> "The Electronic Health Record (EHR) is a longitudinal electronic record of patient health information generated by one or more encounters in any care delivery setting. Included in this information are patient demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data and radiology reports. The EHR automates and streamlines the clinician's workflow. The EHR has the ability to generate a complete record of a clinical patient encounter – as well as supporting other care-related activities directly or indirectly via interface – including evidence-based decision support, quality management, and outcomes reporting." [28].

EHRs are made for primary use, i.e., *meaningful use* for patient's treatment, with an implied trusted domain and confidentiality among medical staff. EHR are also increasingly being used for secondary purposes, such as release of data for governmental health programs and research [20].

2.2.2 Personal Health Record (PHR)

Personal Health Records (PHR) is not as widespread as EHRs. A PHR is a user-centered application that allows individuals to manage their own health information and to share it with other people and/or healthcare providers [54]. PHR can be helpful for maintaining health (fitness and wellness reasons) as well as a tool to help with illness (treatment of patients). Examples of

commercial PHR are HealthVault[1] and PatientsLikeMe[2]. Such systems can be also integrated to other eHealth systems but specially mHealth applications. For instance, heart/glucose monitoring devices and mobile applications (e.g., run/walk trackers, calorie counters).

2.2.3 Mobile Health (mHealth)

Mobile health (mHealth) technology can be defined as the integration of mobile computing, medical sensors, and portable devices to ensure healthcare [31]. The Global Observatory for eHealth (GOe) defined mHealth as:

> "[...] medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices." [56].

mHealth technology is promising in both developed and developing countries. In the developed world mHealth takes advantage of more sophisticated settings, enabling remote monitoring of chronic patients at the comfort of their homes (elderly health and home care) [56]. In developing countries mHealth takes advantage of the flourishing mobile market. Many initiatives use mobile's SMS systems for health campaigns (raising health awareness), treatment adherence, and reminders regarding medication intake [12, 56]. Also, mHealth systems are often used to support frontline health workers, including community health workers (CHWs), nurses and midwifes, in the promotion of primary healthcare [12, 39, 50].

Finally, it is worth noting that E/PHRs and mHealth can be put together, composing fairly complex systems. For instance, a hospital could have an EHR system used by all the medical staff inside its premises. At the same time, many patients already use PHRs to manage and track their own health. Such system could be integrated or have import/export mechanisms for data sharing. Or even, the EHR platform can also be equipped with a patient interface (e.g., a web portal) that allows: access information entered by their physicians; make inputs and rectifications; follow-up of their stages of treatment or see who else has access to their data. Moreover, patient's data coming from mHealth systems can be also shared with E/PHRs. Hospitals can take advantage of homecare systems to remotely monitor patients with chronic conditions (e.g., cardiovascular diseases, asthma, diabetes, cancer, HIV/AIDS). In such case, the patient carries a number of mHealth devices (e.g., mobile phone and sensors) that allow real-time monitoring of their vital signs. Individuals (i.e., patients or app users) may also upload data from their mHealth devices to their E/PHRs.

2.3 Security & Privacy for Healthcare

The respect of privacy is *sine qua non* to healthcare. Its importance, as mentioned in [13], has been already manifested ages ago in one of the most widely known of Greek medical texts, the Hippocratic Oath:

[1] Microsoft HealthVault (`www.healthvault.com`)
[2] PatientsLikeMe (`www.patientslikeme.com`)

> "What I may see or hear in the course of the treatment or even outside of the treatment in regard to the life of men, which on no account one must spread abroad, I will keep to myself, holding such things shameful to be spoken about." – excerpt from Hippocratic Oath [19].

High quality healthcare requires individuals to share their personal health information with healthcare professionals [13]. Furthermore, information should be complete and accurate. If patients cannot trust that their information will be kept secure, they will be reluctant to share it (or even to use the service). If health professionals cannot trust the organization to keep records secure they will not put complete information. In both cases this leads to inferior healthcare. It is therefore paramount that privacy and security concerns are addressed during the design and development of any health information system.

This section introduces the security and privacy technologies related to the thesis. For the sake of clarity, this part of the background is organized in four macro topics: (1) general concerns on security and privacy; (2) security mechanisms; (3) Privacy Impact Assessment (PIA); (4) data obfuscation and anonymisation. That is, moving from the general to the specific concepts.

2.3.1 General Issues of mHealth Security & Privacy

Essentially, mHealth inherits problems from mobile computing and wireless networks. The communication channels are more vulnerable due to their wireless characteristics (e.g., network eavesdropping and spoofing) and mobile devices have more constrained amount of processing power and memory (i.e., need for lightweight cryptography). Devices can be shared among users, and they are more vulnerable to theft, loss and damage, which may result in data breaches, data loss, and privacy infringements.

Regarding general issues linked to mHealth security and privacy, some interesting publications can be discussed. From a more technical perspective, in [38] the authors proposed a number of security and privacy *recommendations* for mHealth developers. These recommendations were made based on a preliminary survey of 169 papers, resulting in the nine general recommendations listed below:

- Access control – Use of patient-centered access control mechanisms (e.g., role-based access control), in which users should be able to allow or deny access to their information at any moment.
- Authentication – Users should be able to authenticate with a unique ID and password (or multi-factor authentication). Passwords should be kept in secrecy and should reach appropriate levels of security.
- Security and confidentiality – Use of encryption mechanisms (e.g., AES) with proper parameter configurations (i.e., key size).
- Integrity – Use of message authentication codes and digital signatures.

- Inform patients – Present privacy policy to users before collection of data that informs about the user rights and specifies the purposes of data collection and processing.
- Data transfer – Use secure communication channels (e.g., TLS, VPNs) while transferring data among entities. Notify user about data transfers.
- Data retention – Inform users about retention policy. The data should be kept only for the necessary time to accomplish the initial purpose. User should be able to check when data is deleted.
- Body Area Network communication – Use security mechanisms for authentication and key distribution among sensors and smart-phones; establish secure communication channels among devices.
- Breach notification – In case of data breaches, the competent authorities and users should be notified. Entities should help users in order to relieve the consequences and restore possible damages.

Overall, the recommendations help developers to have a glimpse about privacy and security issues in mHealth. However, they are incomplete if compared to existing legislations on privacy and data protection, and thus, have limited practical use. In the case of European Union (EU), the General Data Protection Regulation (GDPR) [22] is the upcoming regulation for personal data privacy and protection, replacing the EU Data Protection Directive 95/46/EC [1].

Many countries (i.e., separate legal jurisdictions) however do not have specific legislation for data privacy [26]. This does not imply a legal void in the area, but privacy rights might stem from the constitution or consumer rights; and in the case of healthcare, from medical codes of conduct, and so on. Thus, from the legal perspective, some publications help to bridge this gap between privacy and mHealth technologies. For example, [30] presents a list of five guiding principles for mobile privacy in the context of developing countries (that map to principles of the GDPR):

Principle 1 Address Surveillance Risks – Projects should take steps to ensure that user data is secure from third party surveillance, e.g., user discriminatory profiling can be made by mobile operators and government.

Principle 2 Limit Data Collection and Use – Projects should limit data collection to what is absolutely necessary for the project's goal, e.g., by employing access control, data retention policy, and not collect unnecessary data.

Principle 3 Promote and Facilitate Transparency – Projects should be transparent about what data is collected, how it is shared, and how it might be used in the future, e.g., user notifications, data transfer policies, audit trails of others that also have access to the data.

Principle 4 Incorporate User Feedback – Projects should give users the ability to access, amend, and/or delete their data, e.g., create user interfaces, create communication channels to receive feedback from users.

Principle 5 Assume Responsibility – Projects should assume accountability for potential risks and harms incurred via their projects and platforms, e.g., perform risk assessment, plan incident response, notify data breaches.

The content of the recommendations [38] and the guiding principles [30] offer a good starting point for developers and project leaders. However, in practice, security and privacy analysis should be done case-by-case, given the complexity, multiplicity of actors, jurisdictions, and highly culture-specific dimensions of privacy [41].

2.3.2 Security Mechanisms for mHealth

Here we review some fundamental cryptographic mechanisms and protocols used in the research, specifically for MDCSs. In brief, a Key Management Mechanism (KMM) is used to provide Authentication and Key Exchange (AKE) between parties (user's mobile and application server). Authentication protocols and key derivation schemes for MDCSs usually rely on symmetric cryptography, using password authentication. These protocols should also give support for online and offline user authentication. Other mechanisms should cope with confidentiality of stored and in-transit data, by means of encryption schemes for secure storage and transmission. As a result, the security background herein presented forms the building blocks of our solutions.

Authentication and Key Derivation

Authentication of users remains a challenging and crucial element in modern computer security. Even though authentication mechanisms can be based on a combination of factors – i.e., biometrics ("what the user is"), security tokens ("what the user has"), or passwords ("what the user knows") – the most widespread strategy still lies in secret passwords [44]. This happens because password-based authentication is the most well-known, simple, cost effective and efficient method of maintaining a *shared secret* between a human being and a computer system. Furthermore, the advantages of using passwords tend to out shadow the disadvantages, i.e., problems of choosing strong but easy-to-remember passwords. Thus, it is likely that we will see passwords being used for quite some time into the future [47]; by itself or as part of multi-factor authentication schemes.

Password-based systems normally employ Key Derivation Functions (KDFs), cryptographic algorithms that allow the generation of a pseudo-random string of bits from the password itself [48, sec. 2.4]. Typically, the output of a KDF is employed in one of two manners: it can be locally stored in the form of a *token* for future verifications of the password or it can be used as the *secret*

key for data encryption and/or authentication. Whichever the case, such solutions internally employ a one-way function (e.g., hash), so that recovering the password from the KDF's output turns out to be computationally infeasible [17]. Nonetheless, attackers can still use dictionary attacks [48, sec. 8.1] and test many different passwords combinations until a match is found (i.e., brute force). KDFs usually rely on two basic strategies for preventing such brute-force attacks. The first is to purposely raise the cost of every password guess in terms of computational resources, such as: processing time and/or memory usage. The second is to take as input not only the user-memorisable password, but also a sequence of random bits known as *salt* [3] [48, sec. 3.2]. The presence of such random variable thwarts several attacks based on pre-built tables of common passwords, i.e., forces the attacker to create a new table from the scratch for every different *salt*. The *salt* can, thus, be seen as an index into a large set of possible keys derived from the password, that does not have to be memorized by users or kept in secret.

Password-based Remote Authentication and Key Exchange

In principle, KDFs could be used for data delivery: if the local and remote systems share the same password, they could exchange data by revealing to each other the *salt* employed for generating the key that protects such data. However, since this would allow attackers to use the same *salt* in an offline dictionary attack, KDFs are usually employed only for local data storage, establishing a secure channel between the human user and the local system.

Data delivery to remote locations usually employs Password Authenticated Key Exchange (PAKE) protocols. Such schemes allow two or more parties who share a password to authenticate each other and create a secure channel to protect their communication (for example, [6, 4]). To be considered secure, a PAKE solutions must ensure that an unauthorized party (that fully controls the communication channel but does not know the password) is unable to learn the resulting key and is, as much as possible, unable to guess the password using offline brute force attacks.

The Secure Remote Password (SRP) project group [45] describes the following attacker model for PAKE protocols:

(a) attackers have complete knowledge of the protocol;

(b) attackers have access to a large dictionary of commonly used passwords;

(c) attackers can eavesdrop on all communications between client and server;

(d) attackers can intercept, modify, and forge arbitrary messages between client and server; and,

(e) a mutually trusted third party is not available.

Looking briefly into the history of PAKE protocols, the Encrypted Key Exchange (EKE) [6] was probably the first successful proposal. Although several of the published methods were flawed, the surviving and enhanced forms

[3] **Salt** is a random string that is concatenated with a password ($salt||password$), added to the input before being passed as argument to the one-way function.

of EKE have effectively amplified the security of using passwords to establish shared keys for confidential communication and authentication. Other provably-secure PAKE include the schemes described in [8] (which uses the standard model[4]) and in [5] (which uses the random oracle model[5]). This groups of EKE-inspired proposals are commonly referred as EKE family of protocols.

Forward Secrecy Property

The security of computer systems rely on the condition that attackers cannot gain access to an underlying secret [32]. In practice, however, achieving this condition can be challenging. Most strategies used to minimize exposure of the secret keys end-up raising costs and/or affect system's usability (e.g., multi-factor authentication mechanisms). Therefore, we must assume that a sufficiently motivated adversary may succeed in exposing the system's secrets (specially when using PAKE), and we should explicitly deal with such events and elaborate strategies to minimize potential damages.

One approach, is to use (password-based) protocols that have the so-called *forward secrecy* (also called perfect forward security) property [17]. For PAKE schemes, this property can be translated as follows: if the long-term secret information (e.g., the password) is revealed to an attacker, this information cannot be used to obtain ephemeral keys (i.e., "session keys" derived from the long-term secret) from past communications, effectively protecting all information previously exchanged [53]. That is, if the parties participating in the protocol share a long-term secret S and run the protocol r times before S is compromised by an attacker, that attacker is unable to determine the set of ephemeral keys $K_1, \ldots, K_r$ generated prior to this disclosure of S; only the subsequent keys $K_r + i$ where $(i > 0)$ generated using the same S can be compromised by that attacker.

This concept is an integrating part of many modern security solutions, including pseudo-random generators, digital signatures and public-key encryption [32]. It is usually employed for securing data channels for limited/temporal interaction (i.e., keys expire and should change). Nonetheless, it is also possible to employ the forward secrecy concept for securing data storage, avoiding the encryption of large quantities of data with a single secret key (e.g., as done in OpenPGP's [9] e-mail encryption [52]). Whichever the case, the main drawback of applying forward secrecy is that such strategy incurs additional operations and, most likely, a more complex key management/evolving scheme.

[4] Cryptographic systems are commonly based on complexity assumptions, such as the factorization problem, that cannot be solved in polynomial time. This constructions that can be proven secure using only mathematical complexity assumptions are said to be secure in the standard model.

[5] A random oracle is a mathematical abstraction that *"provides a bridge between cryptographic theory and cryptographic practice"* [5], typically used when the cryptographic hash functions in the method cannot be proven to possess the mathematical properties required by the proof. A system that is proven secure when every hash function is replaced by a random oracle is described as being secure in the random oracle model, as opposed to secure in the standard model.

Secure Data Storage

At the time that user and server agree on a common *shared key* (e.g., the ephemeral keys aforementioned) by means of a PAKE protocol, this key can be used to protect the data stored in the mobile phone. This secure storage mechanism should encrypt all the sensitive information that will reside in the mobile storage (e.g., configuration files, user's data) and the in-transit data that is temporarily stored and sent to the server. Hence, encryption assures data confidentiality letting only authorized parties to read data. This mechanism shall use sufficiently lightweight encryption algorithms owing to the device's limited processing and memory capacity. However, as pointed out by [51], encryption carries the risk of making data unavailable due to data transformation, or if anything goes wrong with the key management process. In other words, developers should be aware that the key management adds complexity since at least on the server-side its necessary to store partial values to rebuild users' keys in order to decrypt and to consolidate received data.

2.3.3 Privacy Impact Assessment (PIA)

Privacy is not only personal data protection, as already mentioned, it has a broader dimension and is more complex than security. For someone to understand privacy, it is crucial to comprehend its technical aspects (e.g., user profiles, data flows, data holders), its security implications, and also consider particular and cultural elements around privacy to a given context. Privacy Impact Assessment (PIA) is a pragmatic manner to make such analysis. This method was encouraged or made mandatory by various legal frameworks for privacy and data protection in different regions (e.g., New Zealand, Canada, Australia, Hong Kong, European Union) [10]. While there is no internationally accepted definition for PIA, we consider the following two definitions:

> [PIA is] "a process whereby the potential impacts and implications of proposals that involve potential privacy-invasiveness are surfaced and examined." [10].

Or a more detailed construction:

> "a privacy impact assessment as a methodology for assessing the impacts on privacy of a project, policy, programme, service, product or other initiative and, in consultation with stakeholders, for taking remedial actions as necessary in order to avoid or minimise negative impacts. A PIA is more than a tool: it is a process which should begin at the earliest possible stages, when there are still opportunities to influence the outcome of a project. It is a process that should continue until and even after the project has been deployed." [59].

PIAs can be tailored to a specific technology and application. For instance, RFID PIA Framework [21] which was developed by industry players and

endorsed by the Article 29 Working Party[6] [59]. Such approach creates a *PIA template* that is pertinent for a specific industry sector. A PIA template establishes a four-stage process [21]: (1) a full description of the application and scenario; (2) identification of privacy threats; (3) a proposal of technical and organizational mitigating measures; and, (4) document the resolution (results of the analysis) regarding the application. In the same way, mHealth developers could benefit from PIA templates.

2.3.4 Data Anonymisation and Obfuscation

High quality healthcare requires sharing data [13]. Access control is one of the straightforward strategies to enforce confidentiality of patient's information in health systems. However, we believe that besides the typical binary decision of revealing or not a data value, access control can be further improved with data *obfuscation*, i.e., to lower individual data item accuracy in a systematic, controlled, and statistically rigorous way [3] to guarantee patient's privacy while retaining its usefulness. For instance, instead of revealing the patient's age one can reveal a range of values. Or even, replace the of medical condition or disease by a more general term (e.g., "Human Immunodeficiency Virus (HIV) infection" replaced by "Infectious Disease").

Besides, individuals' health information are also important to create rich statistical databases for researchers and to support public health programs. In such cases, data *anonymisation* should be employed, i.e., to protect privacy by making a number of data transformations so that individuals whom the data describe remain anonymous. The anonymisation process can have variable degrees of robustness [58], depending on how likely is to: 1) single out an individual in the dataset; 2) link records concerning the same individual; or, 3) infer the value of one attribute based on other values. In essence, all these circumstances should be avoided, resulting in an anonymised dataset. Therefore, anonymised data is not considered personal data, so that, data privacy laws would no longer apply.

3 Research Question

This thesis addresses the following research question.

> *How to design secure and privacy-preserving systems for mobile health data collection and processing?*
>
> By *design* we mean the whole system design process of analysis, specification, modeling, implementation, test, deployment and evaluation of a solution. Mobile systems bring various challenges associated to limited

[6]The Article 29 Data Protection Working Party was set up under the Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. It has advisory status and acts independently. (`http://ec.europa.eu/justice/data-protection/article-29/index_en.htm`)

computing capacities, vulnerabilities of wireless communication channels, and human-computer interaction. In the case of mHealth technologies, the concern with privacy and data security becomes paramount, due to the sensitivity of health information. Although there is no one-size-fits-all solution, this thesis helps to close the security and privacy gap in mHealth solutions for health surveys and surveillance, also know as Mobile Data Collection Systems (MDCS). Altogether, the papers herein presented compose a series of small steps towards answering this research question.

4 Research Method

This research can be categorized as applied science: a discipline of science that applies existing scientific knowledge to develop more (knowledge and) practical applications, such as technology and inventions. Technology is of course supported by scientific knowledge, but also *"other organized knowledge to practical tasks by social systems involving people and machines"* [15]. For this reason, the engineering *Design Method* [16, sec. 1.3.2] (or Design Science [43]) was adopted as predominant research strategy, as an alternative approach to the *Scientific Method* [15]. In short, the Design Method comprises eight main steps, that cyclically iterate from one to another every time assumptions should be redefined.

The Design Method starts with the **problem definition**, basically questioning: [*Who*] need(s) [*what*] because [*why*]?. What is the problem? Who has the problem? Why is it important to solve? So, rather than scientific curiosity, the design is actually driven by needs of society [16].

The second step is the **background research**, i.e., the state-of-the-art that includes scientific knowledge, but it also includes devices, components, market and economic conditions [16]. These steps correspond to the problem characterization and observations in the scientific method.

After understanding the problem and existing solutions, the third step is to **specify requirements**, i.e., the characteristics that your solution should meet. Requirements specification is based on what is known from other existing solutions, as well as by consulting users that need it. Again, this step would correspond to the formulation of a hypothesis, or proposing an explanation.

The fourth step is to **propose a solution**. Researchers should brainstorm solutions within their group and choose the one that better satisfies the project goals, meeting all (or most) requirements. The fifth step, the solution is then further **detailed and modeled**, by means of modeling languages, drawings, and so forth. This need of a model is important to predict the behavior of the solution before prototyping [16]. Similarly, the scientific method needs to define the experiment and its procedures.

Once researchers have a clear idea of what to do, they start the sixth step, to **build a prototype**. In the scientific method this is the test of the hypothesis by experimentation. The seventh step is the **test and redesign** of the prototype by multiple iterations. Similar to the evaluation and improvement

steps in the scientific method. And at last, in both methods, researchers should **communicate results**, through technical reports, publications and documentation.

The Design Method was applied throughout the research with varying levels of completeness, as described below:

Paper I is a survey or review paper. A survey deals with the **problem** of identification, analysis and synthesis of the state-of-the-art in and specific area. In this case, to understand the current situation of mHealth initiatives in Brazil. The research is relevant to mHealth developers, academy, industry and government agencies that could benefit from mHealth solutions in similar settings. By carrying out the survey we initiated the **background research** process, which resulted in a well-structured review of the state-of-the-art. The results were **communicated** by means of a scientific publication. (The remaining steps of the Design Method are not required for survey papers.)

Paper II describes a security framework for MDCS, named SecourHealth. The initial **problem** referred to the design of a security framework for a MDCS that was to be implemented in the city of São Paulo (Brazil). Relevant publications were found during our *ad hoc* **background research**, from which solutions could be however improved and/or adapted to our settings. A new solution was **proposed**, **modeled**, **prototyped** and **tested** in order to demonstrate its feasibility. The results were **communicated** by means of a scientific publication.

Paper III describes a georeferenced and secure MDCS, named GeoHealth. The **problem** refers to the design of a MDCS that could support public primary healthcare in the city of São Paulo. A specific **background research** reveled that existing solution could not meet all desired requirements (e.g., health data quality, security, georeferencing). A new MDCS was therefore **proposed**, **modeled**, **prototyped** and **tested** in order to demonstrate its feasibility. The results were **communicated** by means of a scientific publication.

Paper IV outlines a preliminary PIA template for MDCS. The **problem** consists in the creation of a tool, that mHealth developers can use to objectively assess privacy in their projects. The **background research** revealed a great legal and technological gap between privacy and mHealth systems in general. A PIA template for MDCS was therefore **proposed**, but it should be further developed. Partial results were **communicated** by means of a scientific publication.

Paper V outlines a preliminary ontology-based data sharing system (O-DSS) for medical information. The **problem** refers to a solution to transfer and share individuals' health information in a privacy-preserving manner, by exploiting ontology-based obfuscation and anonymisation functions. The specific **background research** shows that although there

exist viable solutions to be used, they should be linked to realistic use cases and adapted accordingly. An O-DSS was therefore **proposed** and exemplified with use cases, yet it should be still further developed. Partial results were **communicated** by means of a scientific publication.

5 Contributions

Overall, this licentiate thesis contributes to the body of knowledge of security and privacy for mHealth systems. It specifically deals with MDCS, helping developers to understand security and privacy issues and to choose appropriate safeguards. The research also corroborates the notion that security and privacy should be seen as processes, accompanying the whole design life-cycle.

This general contribution is reflected in various partial contributions made in Papers I–V, listed as follows.

1. *Analysis of Mobile Health Systems for Developing Countries.* We provide an in-depth analysis about mHealth initiatives in Brazil (Paper I). It helps researchers to understand current front-runners, target users, types of health applications, adopted devices, and security problems in existing proposals. And also helps us to reflect about the importance of mHealth solutions in primary care settings, potential nation-wide projects, business opportunities and potential research areas. In particular, the security gap for mHealth, since many solutions (in Paper I) have shown little or no concern about protecting collected and processed data.

2. *Security Framework for MDCSs.* We propose SecourHealth, a lightweight security framework designed specifically for MDCSs (Paper II). It provides many security services for both stored and in-transit data, coping with scenario constraints such as network delays, lack of connectivity, device sharing, and security-usability trade-offs. Developers of MDCSs can use the framework, and have it integrated into their solutions or used in the design of more secure applications from the start. We also describe how we integrated SecourHealth into the GeoHealth solution (Paper III), and we present performance benchmarks, showing that it is possible to provide strong security for the data while introducing minimal overhead to the collection process.

3. *Design and Deployment of MDCS for Primary Care.* We propose GeoHealth (Paper III), a secure, low-cost and high-impact MDCS for primary care. In this research, we share our experience regarding the design and deployment of MDCSs, providing evidence for healthcare managers that MDCSs can significantly improve the efficiency and quality of the whole process of health surveys and surveillance. Furthermore, GeoHealth stands out from other MDCSs for having strong security features implemented, which were crucial for its large-scale deployment.

4. *Privacy Impact Assessment for MDCSs.* We provide an analysis about privacy and data protection issues in the context of mHealth system. Also,

we preliminary propose a PIA Template for MDCSs (Paper IV) that can be used by developers and project leaders to properly address privacy in their projects.

5. *Obfuscation and Anonymisation of Medical Data.* We discuss how research areas of security, privacy and ontologies can cooperate to create ontology-based obfuscation and anonymisation functions (Paper V). Based on existing solutions, we propose a Ontology-based Data Sharing System (O-DSS) that makes use of such functions.

6 Related Work

This section describes the related work of thesis. The main existing solutions are briefly presented and (when necessary) compared with the ones proposed in Section 7.

6.1 Mobile Health for Developing Countries

General reports about eHealth and mHealth initiatives in low- and middle-income countries were published by different institutions, e.g., from the United Nations Foundation [12], Earth Institute [39], and World Health Organization [56]. Also, country-specific analyses have been performed, as are the cases of India [23] and China [33]. Even though these publications provide an overall perspective about the topic, there was no comprehensive analysis about the Brazilian scenario. Hence, Paper I presents a survey and in-depth analysis of mHealth solutions and initiatives in Brazil, and also, contributing to spread knowledge that was originally published only in Brazilian Portuguese.

6.2 Security for MDCSs

Although much has been published about mHealth security, there are not many papers that address MDCSs specifically. The most prominent contributions were made by a group of researchers at Bergen University [25, 24, 34, 35]. In this series of works, the authors proposed a security framework for MDCS that covers: user and server authentication, secure data storage and communication. Their solution was integrated in open-source MDCS projects, such as openXdata[7] and Open Data Kit (ODK)[8]. However, the differences between their proposal and the SecourHealth framework (Paper II), are: (a) forward secrecy is added to stored data, and (b) the key management for mutual authentication and data exchange is simpler.

6.3 Design and Deployment of MDCSs

In Brazil, the Primary Care Information System (SIAB) receives primary health care data from all regions of the country, creating a rich database for

[7] `http://www.openxdata.org/`
[8] `https://opendatakit.org/`

health-related action planning. Agents on the field are responsible for visiting families and collecting data (using paper forms), all data is then digitized in the health units, in which they keep a local database but also export the data to SIAB (data consolidation at national level). The MDCS improves the quality and efficiency of entire process. Some existing MDCSs include commercial applications such as Easy SIAB [46] and research projects such as Borboleta [14, 18]. However, the interest of such solutions consist almost solely on their ability to replace paper-based forms with digital ones, leading to the following limitations: (1) they lack support for remote data communication, obliging users to synchronize the collected data only when inside the health units; (2) they do not provide strong security mechanisms for protecting the data stored in the device; and (3) they have no feature for dealing with families having no formal address. These are probably among the reasons why, to the best of our knowledge, they have never been broadly adopted in practice. The proposed MDCS (GeoHealth, Paper III) copes with these limitations, by providing: support for offline and online data collection; protection of data in-transit and at rest; georeferenced data (i.e., GPS for geolocation of families). Also, GeoHealth was deployed on a large scale (total of 28,324 families/96,061 people), proving to be a feasible and low-cost solution (approx. monthly cost of USD 0.04 per inhabitant).

6.4 PIA Template for MDCSs

MDCSs that are inherently privacy-invasive, i.e., surveillance of whole communities. If individuals do not trust the system they will not use it. Privacy incidents in relation to health data can have severe implications for both, data subjects (e.g., discrimination) and health professional (e.g., prosecution and dismissal). If institutions fail to address privacy issues they may face lawsuits, fines, embarrassment and damage to reputation. Thus, we came to realize that instead of running a single privacy assessment for a single MDCS (e.g., GeoHealth, Paper III), it would be more valuable to start building a PIA template for MDCSs (Paper IV), from which different developers and project leaders could benefit.

Besides the general PIA literature (e.g., [10, 59]), the RFID PIA Framework [21] served as example for our PIA template for MDCSs. Other important references are the report on "patient privacy in a mobile world" [41] and publications on security and privacy for MDCSs (e.g., [24, 11]) that support our threat analysis and proposal of countermeasures.

6.5 Obfuscation and Anonymisation of Health Data

This thesis also investigates how to use ontologies for handling textual data values (e.g., diseases, medical procedures, drugs) to decrease semantic loss along the obfuscation and anonymisation process. Ontology-based obfuscation and anonymisation can be used in broad range of systems. The research scope (in Paper V) is however limited to eHealth technologies; and relevant to MDCSs

for data anonymisation, when health data is used for secondary purposes (e.g., statistics and research).

This part of the research was inspired by ontology-based proposals for anonymisation of EHRs [36, 37], and data obfuscation based on users context [57]. In addition, we exemplified how to use obfuscation functions in the Peer Manager[9] [27] by defining them as obligations in the privacy policy. This mechanism could be implemented in health systems (e.g., EHR/PHR) using real medical ontologies (e.g., SNOMED-CT[10]).

7 Summary of Appended Papers

Paper I – Mobile Health in Emerging Countries: A Survey of Research Initiatives in Brazil

Mobile health (mHealth) consists basically in the application of mobile devices and communication capabilities for expanding the coverage and improving the effectiveness of health care programs. This technology is particularly promising for developing countries, in which health authorities can take advantage of the flourishing mobile technology market to bring adequate health care to unserved or underserved communities. Specifically, mHealth can effectively improve basic care and help combating endemic diseases not so often encountered in developed countries. This huge potential has lead to intensive research efforts not only in emerging countries and also around the world, creating a number of innovative solutions. In this paper we provide a comprehensive survey of mHealth research initiatives developed specifically for tackling health challenges in Brazil, an emerging country with a flourishing mobile market. This study identifies the main providers of solution, the areas of deployment, the health conditions that are focus of attention, the types of devices used, the target users, the (lack of) attention to data security issues, among others. Our goal is to discuss gaps, opportunities and tendencies observed in the country, giving some insight on the challenges faced by the mHealth technology in similar scenarios.

Paper II – SecourHealth: A Delay-Tolerant Security Framework for Mobile Health Data Collection

Security is one of the most imperative requirements for the success of systems that deal with highly sensitive data, such as medical information. However, many existing mobile health solutions focused on collecting patients' data at their homes that do not include security among their main requirements. Aiming to tackle this issue, this paper presents SecourHealth, a lightweight

[9] The Peer Manager works as an user-centered identity management platform that keeps user's information private. This framework was built upon the privacy policy language PPL (PrimeLife Policy Language), with which every user can control his personal information by imposing access and usage control restrictions. The Peer Manager is part of the SmartSociety research project (`http://smart-society-project.eu/`).

[10] Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT).

security framework focused on highly sensitive data collection applications. SecourHealth provides many security services for both stored and in-transit data, displaying interesting features such as tolerance to lack of connectivity (a common issue when promoting health in remote locations) and the ability to protect data even if the device is lost/stolen or shared by different data collection agents. Together with the system's description and analysis, we also show how SecourHealth can be integrated into a real data collection solution currently deployed in the city of Sao Paulo, Brazil.

Paper III – Georeferenced and Secure Mobile Health System for Large Scale Data Collection in Primary Care

The Primary Care Information System (SIAB) concentrates basic health care information from all regions of Brazil, providing a rich database for health-related action planning. This data is collected by Family Health Teams (FHTs) in periodical visits to enrolled families in targeted areas. The fact that this procedure relies on paper forms, however, degrades the quality of the information provided to health care authorities and slows down the process of decision making. Aiming to overcome such issues, this article describes GeoHealth, a data gathering application that allows FHTs to use a 3G- and GPS-enabled smartphone for collecting the families' data. Besides quick data validation and delivery, GeoHealth provides strong security features and allows more data to be collected (e.g., the precise location of families having no formal address and extra fields not present in standardized paper forms). We discuss the system's deployment at 6 primary care units in the city of Sao Paulo, where a total of 33.675 families are regularly surveyed. The results obtained show that the process is a low-cost and interesting approach for primary care data collection and analysis.

Paper IV – Towards a Privacy Impact Assessment Template for Mobile Health Data Collection Systems

Mobile Health (mHealth) refers to the use of mobile devices to support health care. Such technologies emerged, especially in developing countries, taking advantage of the flourishing mobile market. Many of them, however, do not properly address the privacy and data protection issues inherent to medical applications. For this reason, aiming to facilitate the developers' work on implementing privacy, this paper motivates and preliminary proposes a Privacy Impact Assessment (PIA) template for Mobile Health Data Collection System (MDCS). PIA templates work as a guiding tool, allowing developers to: (a) understand important privacy principles, (b) identify the privacy threats in their MDCS, and (c) properly mitigate the privacy threats with proper use of technical and administrative controls. Ultimately, this research also intends to foster the development of relevant privacy frameworks for mHealth in general.

Paper V – Ontology-based Obfuscation and Anonymisation for Privacy: A Case Study on Healthcare

Healthcare Information Systems typically fall into the group of systems in which the need of data sharing conflicts with the privacy. A myriad of these systems have to, however, constantly communicate among each other. One of the ways to address the dilemma between data sharing and privacy is to use data obfuscation by lowering data accuracy to guarantee patient's privacy while retaining its usefulness. Even though many obfuscation methods are able to handle numerical values, the obfuscation of non-numerical values (e.g., textual information) is not as trivial, yet extremely important to preserve data utility along the process. In this paper, we preliminary investigate how to exploit ontologies to create obfuscation mechanism for releasing personal and electronic health records (PHR and EHR) to selected audiences with different degrees of obfuscation. Data minimisation and access control should be supported to enforce different actors, e.g., doctors, nurses and managers, will get access to no more information than needed for their tasks. Besides that, ontology-based obfuscation can also be used for the particular case of data anonymisation. In such case, the obfuscation has to comply with a specific criteria to provide anonymity, so that the data set could be safely released. This research contributes to: state the problems in the area; review related privacy and data protection legal requirements; discuss ontology-based obfuscation and anonymisation methods; and define relevant healthcare use cases. As a result, we present the early concept of our Ontology-based Data Sharing Service (O-DSS) that enforces patient's privacy by means of obfuscation and anonymisation functions.

8 Conclusions and Future Work

Health Informatics advance just as far as the individuals' trust in it. Various applications have been created with the hope to further develop healthcare as medical or public health practice. Social development, however, is only real if in consonance with fundamental human rights. Above all, is the right to freedom. Privacy, in turn, has precedence in the right to freedom of opinion and expression, which includes freedom to hold opinions without interference [2]. Without respecting such conditions, new technologies in HI – and other fields – will not achieve their full potential, and may instead be harmful to its users.

In this thesis we deal with the concepts of security and privacy as fundamental principles to achieve high quality healthcare. The research has special focus on mHealth technologies, that have been particularly successful in developing countries. We started with a comprehensive literature review about mHealth initiatives in Brazil (Paper I). Among the various categories of mHealth applications, the class of MDCSs attracted most attention, since such systems handle large amounts of data, for surveillance of whole communities. This, along with the vexing lack of security in existing solutions, provoked us

to continue research in the topic.

In this way, we conducted three additional investigations, aiming (1) to design a security framework for MDCSs (SecourHealth, Paper II); (2) to design MDCS (GeoHealth, Paper III) and share our deployment experience in the city of São Paulo; (3) to propose a PIA template for MDCSs (Paper IV).

As a result, we learned valuable lessons regarding the design and utilization of MDCSs, especially about its users. Community Health Workers (CHWs) play a crucial role in the deliver of basic health and medical care worldwide [50]. They are the first and often the only link between the community and the health system. It is crucial to understand how CHWs make use of mHealth technologies in their work environment in order to design systems that fit their needs. MDCSs should be primarily made for empowering them.

Besides, specially regarding security and privacy, it is important to raise awareness among stakeholders, including the CHWs, project managers, developers, nurses, doctors, and other users. Privacy by design and by default, well-known in the literature and data protection laws, should be put in practice. Health professionals should be conscious that privacy and data protection is a part of their job; and, that its non-observance leads to inferior or inadmissible level of healthcare. In this regard, this thesis presents our implementation experience on secure and privacy-preserving MDCSs.

The privacy aspects of MDCSs should be nevertheless further investigated. Therefore, for future work, we plan to continue the research already introduced in Papers IV and V. Respectively, the completion of the PIA template for MDCSs and additional investigations on medical data obfuscation and anonymisation. Thus, heading towards the design of relevant security and privacy frameworks for MDCS, as well as for mHealth systems in general.

References

[1] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Official Journal L No.281, 23 Nov 1995.

[2] UN General Assembly. Universal declaration of human rights. *UN General Assembly*, 1948.

[3] David E. Bakken, Rupa Parameswaran, Douglas M. Blough, Andy A. Franz, and Ty J. Palmer. Data obfuscation: Anonymity and desensitization of usable data sets. *IEEE Security & Privacy*, (6):34–41, 2004.

[4] Mihir Bellare, David Pointcheval, and Phillip Rogaway. Authenticated key exchange secure against dictionary attacks. In Bart Preneel, editor, *Advances in Cryptology EUROCRYPT 2000*, volume 1807 of *Lecture Notes in Computer Science*, pages 139–155. Springer Berlin Heidelberg, 2000.

[5] Mihir Bellare and Phillip Rogaway. Random oracles are practical: A paradigm for designing efficient protocols. In *First ACM Conference on Computer and Communications Security (CCS)*, pages 62–73, 1993.

[6] Steven M. Bellovin and Michael Merritt. Encrypted key exchange: password-based protocols secure against dictionary attacks. In *Research in Security and Privacy, 1992. Proceedings., 1992 IEEE Computer Society Symposium on*, pages 72–84, 1992.

[7] Matt Bishop. *Introduction to Computer Security*. Addison-Wesley, 2005.

[8] Victor Boyko, Philip MacKenzie, and Sarvar Patel. Provably secure password-authenticated key exchange using diffie-hellman. In *Proceedings of the 19th international conference on Theory and application of cryptographic techniques*, EUROCRYPT'00, pages 156–171, Berlin, Heidelberg, 2000. Springer-Verlag.

[9] Ian Brown, Adam Back, and Ben Laurie. Forward secrecy extensions for openpgp. Technical report, Internet Engineering Task Force, 2001.

[10] Roger Clarke. Privacy impact assessment: Its origins and development. *Computer Law & Security Review*, 25(2):123 – 135, 2009.

[11] Camille Cobb, Samuel Sudar, Nicholas Reiter, Richard Anderson, Franziska Roesner, and Tadayoshi Kohno. Computer security for data collection technologies. In *Proceedings of the Eighth International Conference on Information and Communication Technologies and Development*, page 2. ACM, 2016.

[12] Vital Wave Consulting. mhealth for development: The opportunity of mobile technology for health care in the developing world. Technical report, United Nations Foundation – Vodafone Foundation Partnership, 2009.

[13] Ted Cooper. Managing information privacy & security in healthcare: Privacy and security principles. `http://s3.amazonaws.com/rdcms-himss/files/production/public/HIMSSorg/Content/files/CPRIToolkit/version6/v7/D02_Privacy_and_Security_Principles.pdf`, 2007. Healthcare Information and Management Systems Society.

[14] Rafael Correia, Fábio Kon, and Rubens Kon. Borboleta: a mobile telehealth system for primary homecare. In *ACM Symposium on Applied Computing (SAC'08)*, pages 1343–1347, New York/NY, USA, 2008. ACM.

[15] Nigel Cross, John Naughton, and David Walker. Design method and scientific method. *Design studies*, 2(4):195–201, 1981.

[16] George Ellwood Dieter and Linda C Schmidt. *Engineering design*, volume 3. McGraw-Hill New York, 2013.

[17] Whitfield Diffie, Paul C. Van Oorschot, and Michael J. Wiener. Authentication and authenticated key exchanges. *Des. Codes Cryptography*, 2(2):107–125, June 1992.

[18] Gustavo Duarte, Rafael Correia, Pedro Leal, Helves Domingues, Fábio Kon, Rubens Kon, and Jo ao Eduardo Ferreira. Borboleta and SaguiSaúde – open source mobile telehealth for public home healthcare. In *Proc. of the 8th International eHealth, Telemedicine and Health ICT Forum (Med-e-Tel)*, 2010.

[19] Ludwig Edelstein. *The Hippocratic Oath, Text, Translation and Interpretation*. Baltimore, the Johns Hopkins Press, 1943.

[20] Khaled El Emam. Methods for the de-identification of electronic health records for genomic research. *Genome Medicine*, 3(4):25, 2011.

[21] EU Commission. Privacy and data protection impact assessment framework for rfid applications, 2011.

[22] EU Commission. Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), 2015.

[23] Krishnan Ganapathy and Aditi Ravindra. mhealth: A potential tool for health care delivery in India. *Making the eHealth Connection*, 2008.

[24] Samson Hussien Gejibo. *Towards a Secure Framework for mHealth. A Case Study in Mobile Data Collection Systems*. PhD thesis, The University of Bergen, 2015.

[25] Samson Hussien Gejibo, Federico Mancini, Khalid A. Mughal, Remi A. B. Valvik, and Jørn Klungsøyr. Secure data storage for mobile data collection systems. In *Proceedings of the International Conference on Management of Emergent Digital EcoSystems*, MEDES '12, pages 131–144, New York, NY, USA, 2012. ACM.

[26] Graham Greenleaf. Global data privacy laws: 89 countries, and accelerating. *Privacy Laws & Business International Report*, (115), 2012.

[27] Mark Hartswood, Marina Jirotka, Ronald Chenu-Abente, Alethia Hume, Fausto Giunchiglia, Leonardo A. Martucci, and Simone Fischer-Hübner. Privacy for peer profiling in collective adaptive systems. In Jan Camenisch, Simone Fischer-Hübner, and Marit Hansen, editors, *Privacy and Identity Management for the Future Internet in the Age of Globalisation*, volume 457 of *IFIP Advances in Information and Communication Technology*, pages 237–252. Springer International Publishing, 2015.

[28] HIMSS. Electronic Health Records. `http://www.himss.org/library/ehr/`, 2016. Healthcare Information and Management Systems Society.

[29] HSRIC. Health Informatics. https://www.nlm.nih.gov/hsrinfo/informatics.html, 2016. Health Service Research Information Central.

[30] Hibah Hussain. Dialing down risks: Mobile privacy and information security in global development projects. Technical report, New America Foundation, 2013. [Online; accessed 10-August-2016].

[31] Robert S.H. Istepanian, Emil Jovanov, and Yuan-ting Zhang. Guest editorial introduction to the special section on m-health: Beyond seamless mobility and global wireless health-care connectivity. In *IEEE Transactions on Information Technology in Biomedicine*, volume 8, pages 405–414, 2004.

[32] Gene Itkis. Forward security, adaptive cryptography: Time evolution, 2004.

[33] Ian Leslie, Simon Sherrington, Danny Dicks, Nick Gray, and Tao-Tao Chang. Mobile communications for medical care - a study of current and future health care and health promotion applications, and their use in china and elsewhere. Technical report, University of Cambridge and China Mobile, 2011.

[34] Federico Mancini, Samson Gejibo, Khalid A. Mughal, Remi A.B. Valvik, and Jørn Klungsøyr. Secure mobile data collection systems for low-budget settings. In *Availability, Reliability and Security (ARES), 2012 Seventh International Conference on*, pages 196–205, Aug 2012.

[35] Federico Mancini, Khalid A. Mughal, Samson H. Gejibo, and Jørn Klungsøyr. Adding security to mobile data collection. In *e-Health Networking Applications and Services (Healthcom), 2011 13th IEEE International Conference on*, pages 86–89, June 2011.

[36] Sergio Martínez, David Sánchez, and Aïda Valls. Ontology-based anonymization of categorical values. In *Proc. of the 7^{th} Int. Conf. on Modeling Decisions for Artificial Intelligence (MDAI)*, volume 6408 of *LNCS*, pages 243–254. Springer, Oct. 2010.

[37] Sergio Martínez, David Sánchez, and Aïda Valls. A semantic framework to protect the privacy of electronic health records with non-numerical attributes. *J. of Biomedical Informatics*, 46(2):294–303, 2013.

[38] 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):1–8, 2014.

[39] Patricia Mechael, Hima Batavia, Nadi Kaonga, Sarah Searle, Ada Kwan, Adina Goldberger, Lin Fu, and James Ossman. Barriers and gaps affecting mHealth in low and middle income countries: A policy white paper. Technical report, Center for Global Health and Economic Development Earth Institute, Columbia University, 2010.

[40] Tobias Mettler and Dimitri Aristotle Raptis. What constitutes the field of health information systems? fostering a systematic framework and research agenda. *Health Informatics Journal*, 18(2):147–156, 2012.

[41] mHA. Patient privacy in a mobile world: A framework to address privacy law issues in mobile health. Technical report, mHealth Alliance, 2013. [Online; accessed 10-August-2016].

[42] Claudia Pagliari, David Sloan, Peter Gregor, Frank Sullivan, Don Detmer, P. James Kahan, Wija Oortwijn, and Steve MacGillivray. What is ehealth (4): A scoping exercise to map the field. *J Med Internet Res*, 7(1):e9, Mar 2005.

[43] Ken Peffers, Tuure Tuunanen, Marcus A. Rothenberger, and Samir Chatterjee. A design science research methodology for information systems research. *Journal of management information systems*, 24(3):45–77, 2007.

[44] David Pointcheval and Sebastien Zimmer. Multi-factor authenticated key exchange. In StevenM. Bellovin, Rosario Gennaro, Angelos Keromytis, and Moti Yung, editors, *Applied Cryptography and Network Security*, volume 5037 of *Lecture Notes in Computer Science*, pages 277–295. Springer Berlin Heidelberg, 2008.

[45] SRP Project. What is srp? The Stanford SRP Homepage. Web. 09 Jul. 2013, 2013.

[46] PRSystems. Easy SIAB (in Portuguese). `http://www.siabfacil.com.br/mobile.php`, 2011.

[47] Tom Roeder. Something you know, have, or are. Web. 09 Jul. 2013., 2013.

[48] Bruce Schneier. *Applied cryptography: protocols, algorithms, and source code in C*. Wiley, 1996.

[49] Marcos A. Simplício, Leonardo H. Iwaya, Bruno M. Barros, Tereza C. M. B. Carvalho, and Mats Näslund. Secourhealth: A delay-tolerant security framework for mobile health data collection. *IEEE Journal of Biomedical and Health Informatics*, 19(2):761–772, March 2015.

[50] Prabhjot Singh and Sarah Sullivan. One million community health workers: technical task force report. Technical report, Earth Institute at Columbia University, 2011.

[51] SNIA. Encryption of data at-rest step-by-step checklist. Technical report, Storage Networking Industry Association, 2009.

[52] Hung-Min Sun, Bin-Tsan Hsieh, and Hsin-Jia Hwang. Secure e-mail protocols providing perfect forward secrecy. *Communications Letters, IEEE*, 9(1):58–60, 2005.

[53] Hung-Min Sun and Her-Tyan Yeh. Password-based authentication and key distribution protocols with perfect forward secrecy. *Journal of Computer and System Sciences*, 72(6):1002 – 1011, 2006.

[54] Paul C. Tang, Joan S. Ash, David W. Bates, J. Marc Overhage, and Daniel Z. Sands. Personal health records: Definitions, benefits, and strategies for overcoming barriers to adoption. *Journal of the American Medical Informatics Association*, 13(2):121–126, 2006.

[55] Alan Furman Westin. *Privacy and Freedom*. Atheneum, 1967.

[56] WHO. mhealth: new horizons for health through mobile technologies: second global survey on ehealth. Technical report, The World Health Organization (WHO), 2011.

[57] Ryan Wishart, Karen Henricksen, and Jadwiga Indulska. Context obfuscation for privacy via ontological descriptions. In *Location-and Context-Awareness*, pages 276–288. Springer, 2005.

[58] WP29. Opinion 05/2014 on anonymisation techniques. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf, April 2014.

[59] David Wright. The state of the art in privacy impact assessment. *Computer Law & Security Review*, 28(1):54–61, 2012.

Secure and Privacy-aware Data Collection and Processing in Mobile Health Systems

Information security and privacy are paramount to achieve high quality healthcare services, and further, to not harm individuals when providing care. With that in mind, we give special attention to the category of Mobile Health (mHealth) systems. That is, the use of mobile devices (e.g., mobile phones, sensors, PDAs) to support medical and public health. Such systems, have been particularly successful in developing countries, taking advantage of the flourishing mobile market and the need to expand the coverage of primary healthcare programs. Many mHealth initiatives, however, fail to address security and privacy issues. This, coupled with the lack of specific legislation for privacy and data protection in these countries, increases the risk of harm to individuals. The overall objective of this thesis is to enhance knowledge regarding the design of security and privacy technologies for mHealth systems. In particular, we deal with mHealth Data Collection Systems (MDCSs), which consists of mobile devices for collecting and reporting health-related data, replacing paper-based approaches for health surveys and surveillance.

ISBN 978-91-7063-730-8

ISSN 1403-8099

LICENTIATE THESIS | Karlstad University Studies | 2016:47