Privacy-Preserving Collection and Retrieval of Medical Wearables Data

Christian Maulany Majid Nateghizad Zekeriya Erkin
Delft University of Technology
Cyber Security Group, Department of Intelligent Systems
Mekelweg 4, 2628CD Delft, The Netherlands
c.a.maulany@student.tudelft.nl

Abstract

In recent years we have seen a rise in the amount of fitness tracking and self monitoring devices. These devices which often work in conjunction with a smartphone are becoming more accurate and are becoming widely adopted. This trend goes hand in hand with Electronic Health Care (e-health): the shift of health care to the digital domain. E-health would allow patients to measure their medical condition at home, allowing a diagnosis to be made based on measurements taken over a longer period of time, while reducing the work performed by a doctor. Measurements are stored in the cloud, simplifying the way in which they can be shared with healthcare providers and possibly research institutions. Modernizing healthcare this way should give the patient more insight and control over his/her healthcare and medical data. Furthermore the amount of visits required to the hospital can be reduced, an endeavor which can be demanding for many less fit for elderly individuals.

However, handling medical data this way causes concern for privacy. Often the data handled by these devices is very sensitive and could easily be used to identify the user and monitor many of their behaviours. In order to achieve privacy there are several approaches. One way is to enforce involved parties through legislation to use the data for specific purposes only. However, this relies on the party being semi-trusted and does not guarantee safety in case of a data-breach.

In this work the way in which the integration of wearables into the medical domain is currently taking place and how privacy and security is handled will be explored. Furthermore we will show the current state of research regarding improving this security.

Keywords— E-Health, Wearables, Medical Home Devices, Security and Privacy, End-to-End Encryption, Secure Computation, Encrypted Search

1 Introduction

In 2016 the CBS released a report showing that the percentage of the GDP spend on health care in the Netherlands has risen from 10% in 2000 to 14% in 2015[1]. Furthermore we found the percentage of our population which is older than 65 has risen from 13.57% to 17.80% in this period[2]. This data shows an increasing trend in the cost of health care and an increase in the health care needed for the elderly specifically.

In addition the amount fitness and health tracking devices being sold is increasing, with the market for wearables expected to grow from 84 million units in 2015 to 245 million units in 2019[3]. The accuracy of such sensors and their applications is increasing, while their size and power consumption decreases. We see that many of these

devices are targeted towards consumer use. These devices can be worn throughout the day and are able to measure heartbeat, movement and sleeping patterns. The data gathered by these devices is mainly used for health and fitness tracking. However, we are also starting to see more medically targeted devices, able to measure blood pressure, ECG and glucose levels. These are not necessarily worn and measurement often need to be performed manually. What both types of devices do have in common is that they usually are able to connect to a smartphone. This allows easy cloud storage and analytics on the data, provided by the vendors of these devices.

Fully integrating wearables and health monitoring devices into health care would change health care as we know it, possibly saving costs while at the same time improving the quality of the provided health care[4]. These devices can be used to let patients measure their medical condition at home, reducing the work performed by a doctor reducing costs. Furthermore this approach can reduce the amount of visits required to the hospital, an endeavor which can be demanding for many less fit elderly. Measurements can be taken without the help of a doctor and can easily be taken over a larger period of time. This would possibly allow doctors to make a better diagnosis, or even allow (partly) automation of the diagnosis process. However, in order for such a scenario to be turned into reality first a big step needs to be made into digitalizing existing health care.

At the moment there are several ways in which health care providers (HCPs) manage their data. Traditionally a printed file is kept for every patient, however, most HCPs have moved to local digital file systems. Besides keeping track of the patient's treatment the files are also important when a patient moves to a different HCP, or is visiting multiple providers at the same time. Different types of medications can have different effects when used in conjunction with other medicine. Because of this a HCP needs to be aware of the medication prescribed by the provider itself, but also possibly other providers. For this exchanging patient files between the different HCPs is obviously useful, but is not always performed.

As these smart devices are relatively new, laws and regulations regarding the data they gather often leaves privacy risks. In the United States the Health Insurance Portability and Accountability Act (HIPAA)[5] does not specify data ownership[6] and state laws are inconsistent. We see many vendors of smartphones and smart devices providing their own cloud service for storing and managing the data gathered. Such services include Apple's HealthKit, Gooe Fit, Philip's Health Suite, Witings' Health Mate, and Fitbit's dashboard. As a result data gathered about a single individual exists on many different servers owned by many different parties, each with their own terms of use and privacy policy, leaving little control over the data to the patient. More importantly, ownership of the gathered data often ends up with the service provider instead of the patient. The result is a situation where data which can be considered medical is owned by the service provider who is able to use it for service analytics, targeted advertisements, or even sell it to third parties.

With this the privacy and security challenges of integrating wearables and smart devices into health care should be clear. In this work an overview of the current state of e-health will be given together with suggestions for improving the privacy and security of the current system.

2 Dutch E-Health Architecture

In this section we shall explain the current state of digital health care in the Netherlands.

A system where medical data gathered by HCPs and wearables can easily be aggregated, combined, and exchanged would be a first step into achieving the modernization of health care. In the Netherlands this is done through the Dutch Electronisch Patinten Dossier (EPD) and the Landelijk Schakelpunt (LSP), which translates to electronic

patient file and national exchange point respectively. The EPD was an effort launched by the Dutch government and would become a national system where medical data could be stored. This way it would become easier for patients to look into their own medical files, while at the same time granting the patient more control over who has access to their data and simplifying the way in which medical data could be shared. As a direct consequence medical failure due to misinformation would be reduced.

Preparations for this project started in 2008, however, already in 2011 work on the project halted due to security and privacy concerns. The parliament was not willing to spend the money needed in order to guarantee the system's security. Though parts of the system are still in use, the government would no longer be involved in its development. The result is a decentralized system, meaning that all HCPs participating in the system still store their medical data in their own systems. The LSP provides a centralized point through which HCPs can exchange medical data. All HCPs require an electronic passport allowing identification, while LSP checks the legitimacy of any data requests. Between 90 and 98% of HCPs have moved to a digital file system connected to the LSP[7]. The LSP was originally funded by the Dutch government, now however funding comes from the HCPs paying a fee to join the LSP, which in turn is covered by health insurance companies. Important to note here is that these health insurance companies do not have access to any data collected by the LSP. The LSP stores no actual data, but instead stores the BSN (Unique identifier handed out to each Dutch civilian by the Dutch government) of a patient and which HCPs are allowed to share his data. Furthermore it knows which types of data each HCP is collecting about a patient, but not the actual content.

At the moment this standard for exchanging data in between HCPs is starting to get widely adopted. However, in this standard only HCPs are able to add data to a patients digital file. The patient controls which HCPs are allowed to share data, however, can not add any data to their file. Neither is there a standard for vendors to exchange data with HCPs. At the moment the data gathered by the patient is stored on the vendors owned servers. Often vendors allow some form of data sharing between vendors and HCPs, however, this usually has to be done through their services and protocols. Usually this is either in the form of a web-based dashboard or through an API. The LSP only connects different HCPs. Therefore the LSP does not provide a way for wearables and smart devices to exchange data with HCPs. Furthermore the LSP is not designed to support such an integration.

2.1 Stakeholders

In this section we briefly explain the stakeholders, which rights and powers they have in the current setting, and where the system should move towards in the future, according to our vision on e-health.

- Currently the **patient** is able to decide through the LSP which HCPs are able to exchange medical data. The LSP provides information about which types of medical data are being exchanged, however, does not provide a way to view the data itself. By law a HCP needs to provide the patient insight into which data about them is being collected. For this the patient needs to request access to their medical file at each HCP. The information is generally given by providing a printed copy of the data, instead of allowing digital insight. This means that realtime insight is not possible.

  The patient wants to be able to gain more control over their own medical data. This means easier insight into exactly which data is being shared by the HCPs. More control and involvement can be gained by letting the patient collect their own data using wearables.

- ◇ The main goal of the **vendor** or **service provider** is to persuade patients to use their products. Vendors compete on this based on features provided, quality of measurement, and price. More relevant is the cloud service they provide. A vendor might be able to perform certain analytics on the medical data other vendors might not provide.

  The vendor wants to be able to keep providing cloud-based services in order to compete with other vendors. Simplifying the way by which their measurements can be shared with actual HCPs would make their products more attractive for patients.

- ◇ The **HCP** stores medical data about the patient in order to provide medical treatment. This data can be exchanged through the LSP if the patient has given permission for this.

  Gaining access to measurements gathered by wearables could reduce the time required per patient and improve the care they provide.

- ◇ Right now there are vendors and HCPs who provide anonymized data to **research institutions**. It might be possible that while conducting research the institution might learn a certain patient requires medical attention. If the anonymization process could be reversed under these circumstances the research institution could inform the HCP about this patient.

  Currently vendors often share anonymized data with research institutions. Data gathered by HCPs can be used for personal research or be shared with research institutions, in an anonymized way, with consent from the patient[8]. However the LSP provides no way of achieving or monitoring the exchange of this data.

- ◇ The **LSP** keeps track of which HCPs a patient has given permission to exchange data. The system attempts to only allow them to share data relevant for the treatment they provide. The LSP is responsible for logging any data exchange taking place. These logs can be viewed by the patient.

3 Privacy and Security Considerations

Availability might be an issue. A lot of trust is being put into the LSP. The LSP forms the centralized hub which connects the different parties wishing to exchange data. If the LSP is offline no data can be exchanged between parties.

Sensitive Data is being collected by the vendors. Profile information gathered usually includes full name, E-mail address, date of birth, gender, weight, height. Furthermore depending on the wearables and smart devices used different metric information will be gathered. These types of measurements include hearth rate, blood pressure, weight, fat percentage, amount of steps taken, location. Usually every measurement can be combined by a note or description given by the patient. In the future we can only expect more types of metrics to be aggregated by these systems. Also permissions granted by the patient to any HCP will also be linked to the patient's data. Besides the fact that identifiable information is often included, also proposedly anonymized data is often still susceptible to linkage attacks[9]. Furthermore the data can be used to learn a lot about individuals, their behaviours and daily routines.

Data Visibility. At the moment the vendor has insight into all the patients data gathered by their products. Furthermore they grant themselves the right to use this data to improve their own services, but also the right to use and sell the data to third parties for e.g. advertisement purposes. The LSP is able to see all HCPs a patient is attending and which types data they are collecting and exchanging. Furthermore in case of a data breach the information gathered by vendors will be visible in plaintext to any attacker.

Data Retention. Though most vendors offer patients the option to remove certain measurements, often these measurements will only remove the link between the patient and the measurement. Furthermore the link might even still exist in any backups of the data.

Third parties. Many vendors provide ways through which third parties can use the collected data. This third party might have a different privacy policy, rendering guarantees regarding privacy given by the vendor meaningless. The data can be used for any purpose the third party sees fit, such as advertisement or reselling.

Monitoring is mainly handled by the LSP. The LSP makes sure HCPs only exchange relevant information. There is no way for the patinet to verify the information given by the LSP. When a patient requests insight into their file at a HCP, the HCP is the one supplying this data, usually in printed form. This means the HCP is able to tamper and modify the data.

4 Proposed E-Health Solutions

In order to improve the privacy in e-health several solutions have been proposed. In this section different relevant works will be explained briefly.

4.1 Privacy through Encryption

4.1.1 End-to-End Encryption

By storing the data in an encrypted form it is possible to hide it from the vendor. One way to achieve data sharing is by combining encryption at the endpoints with proxy re-encryption[10]. Proxy re-encryption allows encrypted data to be re-encrypted so that it can be decrypted using a different key. This way data can be shared among multiple parties without the need of sharing keys. Furthermore the encryption at the endpoints means that any intermediate party can not view the data. Finally the scheme still allows computations to be performed over the encrypted data, leaving room for an extension where the storage provider provides some diagnostic service over the encrypted data.

4.1.2 Polymorphic Encryption and Pseudonomysation

One way to achieve anonymization is through pseudonomysation [11]. A user's actual identifier gets encrypted. Next the encrypted identifier can get reshuffled so it will decrypt to different pseudonyms of this identifier. By reshuffling to a different pseudonym for every party, identifiers can not be linked across different databases. A separate party takes care of the reshuffling and reshuffles the identifiers so data can be exchanged between different parties.

The basis of this system relies on an asymmetric ElGamal cryptosystem. The cryptosystem uses randomized encryption and allows re-keying, re-shuffling and re-randomization are possible. Re-keying allows a ciphertext to be modified so that afterwards it can be decrypted using a different secret key. Re-randomization will change the randomness in the encryption and re-shuffling will modify the ciphertext so that it will decrypt to a differnet message.

The system achieves hiding of identifiers across different parties. However all data remains visible in plain-text at the endpoints. This still leaves the system vulnerable to linking and re-identification attacks, as exchanged data will be visible in plaintext.

4.2 Secure Computation

Encrypting the data is great for provide cloud-privacy. However often vendors perform some computation on the data in order to provide some service or gain some analytical insight in the usage of their product. Some vendors or service providers allows remote diagnostics of biomedical signals. Such a service could be a standalone service, or part of a complete e-health system providing storage and processing of medical data. The user wants an analysis of their data without loss of privacy. At the same time the service provider does not want to reveal their algorithm, as revealing this would allow anyone to provide the same service.

4.2.1 Garbled Circuits

Garbled Circuits provide a way by which to parties can jointly evaluate a function over their private inputs[12]. The one limitation being that function to be evaluated needs to be described as a Boolean Circuit. Here one party constructs the garbled circuit while the other party performs the evaluation. In the scenario of a patient wanting to obtain a diagnosis without revealing his/her data, and a service provider in possession of an algorithm to provide this diagnosis, garbled circuit can provide a solution. By converting the diagnosis algorithm into a Boolean Circuit, this circuit can be garbled by the vendor and evaluated by the patient without loss of privacy.

4.2.2 Linear Branching Programs

By modeling the service as a Linear Branching Program (LBP) the service can be executed even on encrypted data[13]. The basis is an LBP which can classify ECG signals, thus providing a diagnosis. By evaluating this LBP in a privacy secure manner using two-party computation it is shown that such a system can be implemented in an efficient and secure manner. Two implementations, one using garble circuits and one using a hybrid of homomorphic encryption and garbled circuits (hybrid) are proposed. Both provide an accurate classification in a timely manner, with the garbled circuit implementation running more efficient, but the hybrid implementation requiring less communication. Though this work shows that such tools can be implemented in an effective and efficient manner, they do require a lot of tailoring specific to the application in order to achieve the optimizations.

4.3 Searching over Encrypted Data

Storing the data in encrypted form means that performing search over the data becomes a challenge. Especially the use of wearable might greatly increase the amount of data over which needs to be searched.

One way to provide search over encrypted data is by giving each file a limited set of keywords. Using some one-way function, such as a hash, these keywords are then hidden. The hash gets stored on the server along with a pointer to the associated file. A search can then be performed by comparing the hashes of keywords. This provides a very basic search functionality, the keywords require an exact match otherwise the query will give no result. In order to account for typos fuzzy search can be applied[14][15]. The edit distance between two words is the amount of operations required to turn one word into another. An operation can be the addition, removal or replacement of a character. In this setup a maximum edit distance between a searched keyword and a return keyword is defined. When a keyword is added to the database not only the keyword hash itself, but also all keyword hashes within the maximum edit distance are stored. When searching for a keyword a search is performed for the hash of the keyword itself and all keywords within the edit distance.

5 Discussion

The potential of e-health should be clear. E-health can reduce the workload put on doctors and give more control to patients. This could potentialy save costs and improve the quality of health care. However, the architecture currently in place has its limitations. The LSP provides a way for HCPs to exchange data but not for third parties. Because of this integrating wearable into health care is still non-standardized and difficult. The collected data could be of high value to research institutions. HCPs and vendors decide which researchers gets access to the data, meaning the patient has no insight into who has access to their (anonymized) information. The current architecture has certain privacy and security considerations. The level of trust put in different parties is high. Both the LSP and vendors receive high levels of trust, with the vendors granting themselves ownership over the collected data, sharing it with third parties as they see fit. Because of this it is difficult for patients to monitor for what purposes their data is being used.

The scientific work in cryptography provides several tools to address these challenges. These tools all solve part of the challenges and together could form the basis of a system architecture providing the basis for a modern e-health system where privacy would be preserved. End-to-end encryption can be used to hide data from vendors, while secure computation allows them to still provide the same services. By providing proper methods for searching over this encrypted data an architecture can be designed with privacy by design. More importantly control over the data moves from HCPs and vendors back to the patient.

References

[1] Centraal Bureau voor de Statistiek. Zorguitgaven; kerncijfers. statline.cbs.nl/Statweb/publication/?VW=T&DM=SLNL&PA=83037NED&D1=a&D2=&HD=160517-1200&HDR=G1&STB=T.

[2] Centraal Bureau voor de Statistiek. Bevolking; kerncijfers. http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=37296ned&D1=0-2,8-13,19-21,25-35,52-56,68&D2=0,10,20,30,40,50,60,64-65&HDR=G1&STB=T&VW=T.

[3] CC S Insight. Wearables market to be worth $25 billion by 2019. http://www.ccsinsight.com/press/company-news/2332-wearables-market-to-be-worth-25-billion-by-2019-reveals-ccs-insight.

[4] John A. Stankovic. Wireless sensor networks for home health care. In *PECCS 2013 - Proceedings of the 3rd International Conference on Pervasive Embedded Computing and Communication Systems, Barcelona, Spain, 19-21 February, 2013*, 2013.

[5] U.S. Government Publishing Office. H. rept. 104-736 - health insurance portability and accountability act of 1996, 1996. https://www.gpo.gov/fdsys/search/pagedetails.action?granuleId=CRPT-104hrpt736&packageId=CRPT-104hrpt736.

[6] Medical Economics. Patient records: The struggle for ownership. http://medicaleconomics.modernmedicine.com/medical-economics/news/patient-records-struggle-ownership.

[7] VZVZ. 10 feiten over het lsp. https://www.vzvz.nl/page/Zorgconsument/Links/Over-VZVZ/10-feiten-over-het-LSP.

[8] Ondernemingsdatabank. Patintgegevens voor wetenschappelijk onderzoek? `http://ondernemingsdatabank.indicator.nl/patieentenrecht/patientgegevens_voor_wetenschappelijk_onderzoek_/NLTAMDAR_EU030302/68/search`.

[9] Martin M. Merener. Theoretical results on de-anonymization via linkage attacks. *Trans. Data Privacy*, 5(2):377–402, 2012.

[10] K. Rohloff and Y. Polyakov. An end-to-end security architecture to collect, process and share wearable medical device data. *E-health Networking, Application & Services*, pages 615 – 620, 2015.

[11] Eric R. Verheul, Bart Jacobs, Carlo Meijer, Mireille Hildebrandt, and Joeri de Ruiter. Polymorphic encryption and pseudonymisation for personalised healthcare. *IACR Cryptology ePrint Archive*, 2016:411, 2016.

[12] Riccardo Lazzeretti and Mauro Barni. Private computing with garbled circuits [applications corner]. *IEEE Signal Process. Mag.*, 30(2):123–127, 2013.

[13] Mauro Barni, Pierluigi Failla, Riccardo Lazzeretti, Annika Paus, Ahmad-Reza Sadeghi, Thomas Schneider, and Vladimir Kolesnikov. Efficient privacy-preserving classification of ECG signals. In *First IEEE International Workshop on Information Forensics and Security, WIFS 2009, London, UK, December 6-9, 2009*, pages 91–95, 2009.

[14] Jin Li, Qian Wang, Cong Wang, Ning Cao, Kui Ren, and Wenjing Lou. Fuzzy keyword search over encrypted data in cloud computing. In *INFOCOM 2010. 29th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies, 15-19 March 2010, San Diego, CA, USA*, pages 441–445, 2010.

[15] Cong Wang, Kui Ren, Shucheng Yu, and Karthik Mahendra Raje Urs. Achieving usable and privacy-assured similarity search over outsourced cloud data. In *Proceedings of the IEEE INFOCOM 2012, Orlando, FL, USA, March 25-30, 2012*, pages 451–459, 2012.