Towards Privacy by Design in Personal e-Health Systems

George Drosatos[1], Pavlos S. Efraimidis[2], Garrath Williams[3] and Eleni Kaldoudi[1]
[1]*School of Medicine, Democritus University of Thrace, Dragana, Alexandroupoli, Greece*
[2]*Dept. of Electrical & Computer Engineering, Democritus University of Thrace, Kimmeria, Xanthi, Greece*
[3]*Department of Politics, Philosophy and Religion, Lancaster University, Lancaster, United Kingdom*
{gdrosato, pefraimi}@ee.duth.gr, g.d.williams@lancaster.ac.uk, kaldoudi@med.duth.gr

Keywords: Privacy by Design, Personal e-Health Systems, Privacy-Enhancing Technologies.

Abstract: Personal e-health systems are the next generation of e-health applications and their goal is to assist patients in managing their disease and to help both patients and healthy people maintain behaviours that promote health. To do this, e-health systems collect, process, store and communicate the individual's personal data. This paper presents an analysis of personal e-health systems and identifies privacy issues as a first step towards a 'privacy by design' methodology and practical guidelines.

1 INTRODUCTION

An aging population, increasing rates of chronic diseases, and rising healthcare costs represent important pressures towards forms of self-management of health and disease outside health care institutions. New techniques of self-management have become feasible owing to the advent of a variety of personal e-health systems, including wearable sensors (Swan, 2012), personal health records (Johansen and Henriksen, 2014) and self-management and empowerment applications for a number of diseases (Samoocha et al, 2010), delivered via smart phones or other portable personal devices (Mosa et al, 2012), as well as via integrated smart home environments (Teng et al, 2008; Pantelopoulos and Bourbakis, 2010).

Personal e-health systems are designed to be used by the citizens themselves to acquire, store, and manage personal health data. This single user access makes it easy to forget or ignore the inherent security and privacy risks involved. Privacy-related legislation, e.g. the European Data Protection Directive (European Parliament, 24 Oct. 1995) and the HIPAA (Health Insurance Portability and Accountability Act) (104th U.S. Congress, 21 Aug. 1996) explicitly defines the rules for protecting the privacy of patients and covers issues such as access rights to data, how and when data are stored, security of data transfer, data analysis rights, and governance policies. However, it is widely recognized that taking a strong regulatory approach is not always enough, and that privacy safeguards should be built in the design, operation and management of information processing technologies and systems (European Commission, 2012).

This paper focuses on contemporary personal e-health systems and offers a generic description of their functionalities. Privacy concerns for each modeled system's functionality are discussed and possible technical solutions are summarized. The domain analysis presented here is the first step towards a methodology for engineering privacy in the design of a personal e-health system, and practical guidelines for selecting and developing appropriate privacy preserving techniques.

2 PERSONAL DATA AND PRIVACY

Information privacy refers to the legal right to privacy in the collection and sharing of data about oneself. Privacy concerns arise wherever uniquely identifiable data relating to a person are collected and stored, in digital form or otherwise (European Parliament, 24 Oct. 1995). Privacy is related to, but not to be confused with data security, which refers to protecting data from risk of destruction or alteration and from unauthorized use. Here we focus solely on data privacy.

Drosatos, G., Efraimidis, P., Williams, G. and Kaldoudi, E.
Towards Privacy by Design in Personal e-Health Systems.
In *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2016) - Volume 5: HEALTHINF*, pages 472-477
ISBN: 978-989-758-170-0
Publisher: SCITEPRESS - Science and Technology Publications, Lda

A basic data privacy principle refers to the importance of *individual consent* and *control*: the right of each individual to protect her privacy by retaining control over her personal data and knowing who, when and why gets access to her data. Further principles include those of *data minimisation*, *data protection by design*, and *data protection by default* (European Commission, 2014). *Data minimisation* means that when an authority/party requires some information in order to provide support, only the minimum amount of personal information needed to give that support is transmitted. *Data protection by design* is about engineering privacy measures into each part of a personal data system. Finally, *data protection by default* requires that the default operation of any system should be to preserve privacy.

Primary privacy concerns include (Hansen, 2012): (A) *User identification:* The variety of personal data and the quantity of them constitute a key risk factor in undermining privacy. The greater the amount of personal data an adversary possesses, the better able she will be to identify an individual (Narayanan and Shmatikov, 2010). (B) *Personal data leakage:* Careful management and storage of personal data are crucial, especially when these data are health records (Kierkegaard, 2012). Appropriate security should be applied to avoid accidental disclosures and handle potential attacks. (C) *Linkability issues:* A user may appear in several datasets of institutional systems. Precautions have to be taken to avoid the linkage by an adversary of the corresponding partial profiles of the user, since this would generate a larger and more revealing profile. Technical countermeasures like pseudonyms and data anonymization, and policy measures (clear terms on data usage) can be used against this privacy threat. Thorough treatment of privacy concerns and principles are given elsewhere (Danezis et al, 2015).

3 PRIVACY IN PERSONAL E-HEALTH SYSTEMS

Personal e-health systems are designed to be used by citizens or patients themselves in order to maintain their health and manage disease mainly outside the healthcare context, thus promoting heath literacy, disease prevention, integrated care, and self-management. Apart from traditional health-related personal data, such as health records and biomedical sensors' data, personal e-health systems may also utilize data from the user's surroundings, the user's web activity, and other health-related services.

In general, a personal e-health system acquires, stores and processes personal health data, either manually entered by the individual or collected via other personal systems, e.g. sensors or e-health applications. This might also be complemented by data on the environment of the individual (e.g. geolocation, temperature, allergens, etc.), again usually acquired via personal sensors or the mobile device itself. Furthermore, a personal e-health system may require personal data from medical and

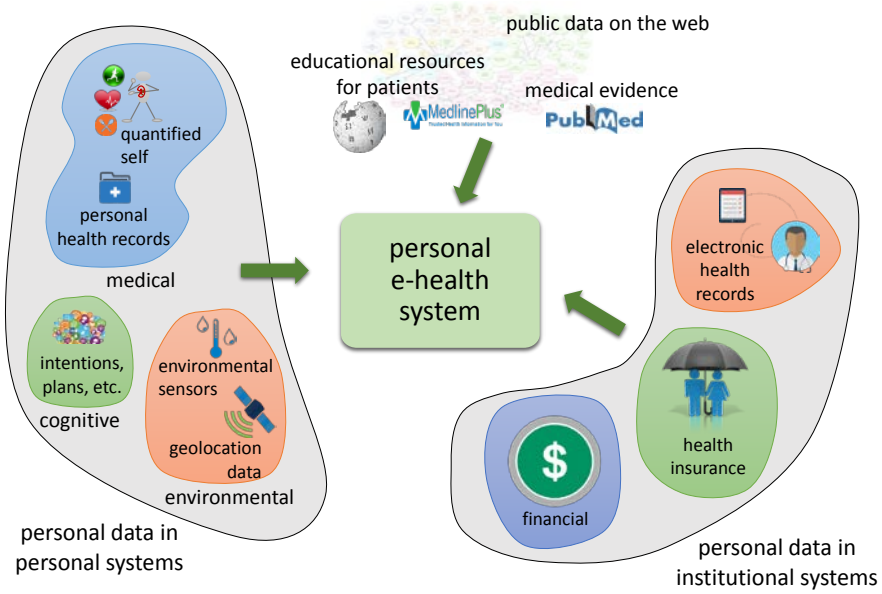

Figure 1: Data communication requirements in a personal e-health system.

other institutional systems, e.g. medical health record segments, electronic prescriptions, insurance and financial details. Finally, personal e-health systems may require access to public databases, e.g. medical ground knowledge or health educational material. Figure 1 presents a graphical overview of data requirements for a personal e-health system.

Based on the requirements for personal data communication, we can identify the following five basic personal e-health systems functionalities (Figure 2): (1) personal data storage and processing; (2) personal data exchange with other third party systems (personal or institutional); (3) integration of (personalized) public data; (4) exporting personal data for public (e.g. statistical) use; (5) exchange of private personal data messages.

3.1 Acquisition, Storage and Processing

Storage and processing of personal data are the core components of a personal e-health system. When both components are located on a user device then privacy can generally be maintained by default. However, nowadays, the most common case is that storage and/or processing are located on a remote server and most often on a cloud infrastructure.

In case personal data are stored on a remote service, their security and privacy need to be ensured. The most common techniques for this are cryptographic techniques and especially techniques that perform client-side encryption of data to protect against untrusted service providers (i.e., Cloud providers). A good review of the cryptographic mechanisms for data storage in the remote services (and especially in the Cloud) is provided in (Kamara and Lauter, 2010) and more general advanced cryptographic schemes are given in chapter 5 of (Smart et al, 2014a).

However, simple encryption of stored data is generally not efficient because at some point some data processing (even a simple search and retrieve) will be required. In such a case the user would have to allow the service provider to decrypt the data (thus compromising privacy), or download all data to the user-side to decrypt and process, or use some computationally intensive approach like searchable encryption. In general, data processing is a complex procedure involving dedicated logical checks, computations and searching over personal data. Thus, there are no generic solutions to support processing of encrypted data. There are some approaches to this problem that offer varying degrees of privacy and/or processing quality assurance, as discussed below.

The most privacy preserving approach is to process encrypted data. There are a number of emerging technologies, such as (fully) homomorphic encryption and searchable encryption, which aim to give general solutions in this direction (Smart et al, 2014b) or even simpler homomorphic techniques that may require some pre-processing (Drosatos and Efraimidis, 2014). However, all these techniques have the following limitations: (a) data should be generally pre-processed before encryption; (b) processing of encrypted data is computationally more intensive than processing of unencrypted data; and (c) all these techniques cannot, in practice, be applied in all cases but have to be considered on a case-by-case basis.

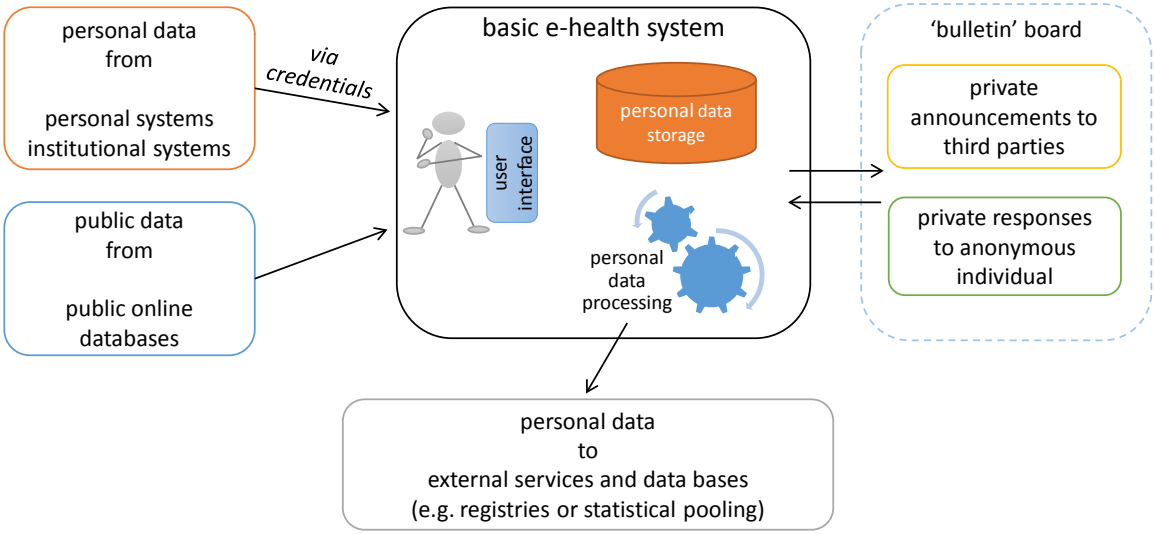

Figure 2: Modelling basic functionalities of a personal e-health system.

An alternative approach involves using a user proxy service to sanitize data, includins such measures as anonymizing, minimizing, transforming and/or aggregating personal data before submitting them for (unencrypted) remote processing (e.g. Layouni et al, 2009). In this approach anonymous credentials (Camenisch and Lysyanskaya, 2001) can be used to prove that the proxy corresponds to a valid system user and at the same time allow anonymity to be revoked under special predefined circumstances (e.g. if a life threatening situation is detected as a result of processing).

From the point of view of data security, personal data should be encrypted as close to their generation as possible, preferably at their source. This imposes additional demands on personal sensor devices commonly used as a data source for a number of personal e-health systems.

3.2 Data Exchange with Other Systems

A quite common requirement or functionality in personal e-health systems is to share and exchange data with other similar systems. For example, personal health record systems usually provide the functionality of integrating data from a number of personal biomedical sensors, such as the free personal health record service HealthVault (HealthVault, 2015) or the personal health avatar service by the MyHealthAvatar project (MyHealthAvatar, 2015). Personal decision support systems may also require integration of other sources of personal data, e.g. the health risk predictive system developed by the CARRE project (CARRE, 2015). Less commonly, data might be exchanged with institutional systems that hold personal data (e.g. personal insurance or financial data or even electronic health records maintained by healthcare providers).

Such personal data exchange between two personal systems requires that one system knows and uses the user's credentials for authentication in the second system. This gives rise to two major problems. The first concerns the potential for malevolent use of the other system's credentials. This actually represents more of a security problem and will not be further discussed here. The second problem concerns linkability, that is, linkage of the different user accounts in various personal systems to a single user. Linkability is a more general concept in personal systems. The most basic linkability relates a system user to an actual physical person. In personal systems this can be more easily achieved, as the system user does not have to be directly linked to a physical person via a strong identifier. For example, in a personal system the user may decide to use pseudonyms (although this may not entirely solve the problem as a person may be identifiable by other data even when her name is not known). When a system knows and uses different user accounts on different systems the use of pseudonyms represents the most usual approach to preserve anonymity and thus, indirectly, privacy. However, integrating partial personal data of the user (as residing in each individual system) to a larger and thus more comprehensive and revealing data set increases privacy concerns.

Generally, there is no direct remedy for this problem. The most obvious solution involves building dedicated middleware that will act as a user proxy for all personal systems. This would reside on the user side and would unlink the flow of personal data among the systems, hiding each system and system account from the other.

3.3 Integration of Public Data

Personal e-health systems may also involve runtime integration of personalized public data. A common example is to fetch on-line publicly available health promotion and educational material suited for the particular user's condition, another example, to fetch information on healthcare resources (nearest doctor, specialty hospital, etc.). Although the data are publicly available, just the act of linking particular data to a specific user may cause a privacy violation, by revealing the user's presumed health care needs.

There are a number of proposed techniques to conceal user requirements by altering the initial request, e.g. by expanding and generalizing the request for public data. These techniques fetch a large amount of data to the user application and then a second round of local processing extracts the specific data relevant to the user (Drosatos et al, 2015). Other emerging approaches require the cooperation of a group of users in the system to conceal one another's requests (Romero-Tris et al, 2015).

An alternative is to use anonymous network technologies that protect the physical address of user from the public service. A representative example is the TOR service (Dingledine et al, 2004), which creates a network of proxies over the internet and allows recursive message encryption along the chain of proxies.

3.4 Exporting Personal Data

Personal e-health systems may need to export anonymised personal data to external services (e.g. medical registries) and/or provide data for statistical use. Exporting personal data to medical registries raises the problem of de-identification (Fung et al, 2010). Here data should be minimized and stripped of all identifiable parts. Some examples of data that can be used to identify an individual include the identity number, or a combination of the birth date and the zip code. However, it is vital to remember that it may be possible to identify a person even when seemingly non-identifiable data are released. One of many interesting example involves the identification of a woman in the United States based on processing of anonymous web search engine query-logs from about 650,000 users over three months (Pass et al, 2006).

When exporting personal data for statistical data pooling, privacy preservation can be promoted by a number of techniques that compute aggregated results (e.g. Lindell and Pinkas, 2009; Drosatos and Efraimidis, 2014). The privacy issues that arise in this type of system depend on the number of patients who are included in an aggregated result – too small a number may still reveal sensitive information about the participants. Another sort of privacy goal involves concealing personal information from the processing module (similarly to a voting system). Here, the selection of appropriate techniques depends on the location (remote or at the user) of storage and the particular form of statistical processing.

3.5 Exchange of Private Data Messages

Occasionally, personal e-health systems may need to exchange private data messages with trusted parties. This includes communicating with a medical professional or a family member. This data communication may be eponymous, that is the user chooses to reveal their identity. However, following the privacy by default principle, the general case must require the anonymous exchange of personal data messages. In this process the receiving party is unaware of the identity of the sender; however, they can still respond and return a data message. This can be achieved using anonymous credential techniques (Camenisch and Lysyanskaya, 2001). Messages can be exchanged via a bulletin board where the original data message and the response are published. The confidentiality of exchanged data messages between the end users (sender and receiver) can be achieved using a secret pre-agreed key for the encryption of messages. A representative example for privately and unlinkably exchanging messages is presented in (Hoepman, 2015).

4 DISCUSSION

This paper focuses on contemporary personal e-health systems and presents a generic high-level model of their functionalities. Privacy concerns for each functionality have been discussed and possible technical solutions presented.

Here we have given only a high-level overview of personal e-health system functionalities; more detailed or even case-by-case analysis would be required to thoroughly cover the plethora of personal e-health applications. Furthermore, the focus of our analysis is on privacy. Data security, while essential, is not discussed as it is generally treated as a lower level storage and communication prerequisite.

Work in progress takes into account the analysis presented here as the basis of a formal step-by-step methodology for building privacy preserving personal e-heath applications. Such a methodology can then be combined with available reviews of privacy strategies and technical solutions (e.g. Danezis et al, 2015) to create a set of practical guidelines for selecting the ideal privacy enhancing technologies in the development of new personal e-health systems.

ACKNOWLEDGEMENTS

This work was supported by the FP7-ICT project CARRE (No. 611140), funded in part by the European Commission.

REFERENCES

104th U.S. Congress, 21 Aug. 1996. Health insurance portability and accountability act. In *Public Law* 104-191.

Camenisch, J., Lysyanskaya, A., 2001. An efficient system for non-transferable anonymous credentials with optional anonymity revocation. In *EUROCRYPT '01, Advances in Cryptology*. Springer Berlin Heidelberg, 93-118.

CARRE Project, accessed 1 Nov. 2015. Personalized patient empowerment and shared decision support for cardiorenal disease and comorbidities. Funded by

European Commission (No. 611140), https://www.carre-project.eu

Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J. H., Metayer, D. L., Tirtea, R., & Schiffner, S., 2015. *Privacy and Data Protection by Design-from policy to engineering*. European Network and Information Security Agency (ENISA).

Dingledine, R., Mathewson, N., Syverson, P., 2004. Tor: the second-generation onion router. In *Proc. of the 13th USENIX security symposium*, 303-320.

Drosatos, G., Efraimidis, P. S., 2014. User-centric privacy-preserving statistical analysis of ubiquitous health monitoring data. *Computer Science and Information Systems*, 11(2), 525-548.

Drosatos, G., Efraimidis, P. S., Arampatzis, A., Stamatelatos, G., Athanasiadis, I. N., 2015. Pythia: A Privacyenhanced Personalized Contextual Suggestion System for Tourism. In *COMPSAC '15, Proc. of the 39th Annual IEEE Computer Software and Applications Conference*. IEEE, 822-827.

European Commission, 2012. 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), http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf.

European Commission, 2014. Green Paper on Mobile Health ("mHealth") (SWD(2014) 135 Final). https://ec.europa.eu/digital-agenda/en/news/summary-report-public-consultation-green-paper-mobile-health.

European Parliament, 24 Oct. 1995. Directive 95/46/EC. In *Official Journal L 281*, 0031-0050.

Fung, B., Wang, K., Chen, R., Yu, P. S., 2010. Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys (CSUR)*, 42(4), 14.

Hansen, M., 2012. Top 10 mistakes in system design from a privacy perspective and privacy protection goals. In *Privacy and Identity Management for Life*. Springer Berlin Heidelberg, 14-31.

Hoepman, J. H., 2015. Privately (and Unlinkably) Exchanging Messages Using a Public Bulletin Board. In *Proc. of the 14th ACM Workshop on Privacy in the Electronic Society*. ACM, 85-94.

Johansen, M. A., & Henriksen, E., 2014. The evolution of personal health records and their role for self-management: A literature review. *Stud Health Technol Inform*, 205:458-462.

Kamara, S., Lauter, K., 2010. Cryptographic cloud storage. In *Financial Cryptography and Data Security*. Springer Berlin Heidelberg, 136-149.

Kierkegaard, P., 2012. Medical data breaches: Notification delayed is notification denied. *Computer Law & Security Review*, 28(2), 163-183.

Layouni, M., Verslype, K., Sandıkkaya, M. T., De Decker, B., Vangheluwe, H., 2009. Privacy-preserving telemonitoring for ehealth. In *Data and Applications Security XXIII*. Springer Berlin Heidelberg, 95-110.

Lindell, Y., Pinkas, B., 2009. Secure multiparty computation for privacy-preserving data mining. *Journal of Privacy and Confidentiality*, 1(1), 59-98.

HealthVault, accessed 1 Nov. 2015. A web-based platform to store and maintain health and fitness information. Microsoft Corporation, https://www.healthvault.com

Mosa, A. S. M., Yoo, I., Sheets, L., 2012. A systematic review of healthcare applications for smartphones. *BMC Med Inform Decis Making*, 12(1), 1-31.

MyHealthAvatar Project, accessed 1 Nov. 2015. A demonstration of 4D digital avatar infrastructure for access of complete patient information. Funded by European Commission (No. 600929), http://www.myhealthavatar.eu

Narayanan, A., Shmatikov, V., 2010. Myths and fallacies of personally identifiable information. *Communications of the ACM*, *53*(6), 24-26.

Pantelopoulos, A., Bourbakis, N. G., 2010. A survey on wearable sensor-based systems for health monitoring and prognosis. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 40(1), 1-12.

Pass, G., Chowdhury, A., Torgeson, C., 2006. A picture of search. In *InfoScale '06, Proc. of the 1st international conference on Scalable information systems*. ACM Press.

Romero-Tris, C., Viejo, A., Castellà-Roca, J., 2015. Multi-party methods for privacy-preserving web search: Survey and contributions. In *Advanced Research in Data Privacy*. Springer International Publishing, 367-387.

Samoocha, D., Bruinvels, D. J., Elbers, N. A., Anema, J. R., van der Beek, A. J., 2010. Effectiveness of web-based interventions on patient empowerment: a systematic review and meta-analysis. *J Med Internet Res*, 12(2).

Smart, N., Rijmen, V., Gierlichs, B., Paterson, K. G., Stam, M., Warinschi, B., Watson, G., 2014a. *Algorithms, key size and parameters report*. European Network and Information Security Agency (ENISA).

Smart, N., Rijmen, V., Stam, M., Warinschi, B., Watson, G., 2014b. *Study on cryptographic protocols*. European Network and Information Security Agency (ENISA), Report TP-06-14-085-EN-N, 11.

Swan, M., 2012. Sensor Mania! The Internet of things, wearable computing, objective metrics, and the quantified self 2.0. *J Sens Actuator Netw*, 1, 217-253.

Teng, X. F., Zhang, Y. T., Poon, C. C., Bonato, P., 2008. Wearable medical systems for p-health. *IEEE Reviews in Biomedical Engineering*, 1, 62-74.