eHealth Beyond the Horizon – Get IT There
S.K. Andersen et al. (Eds.)
IOS Press, 2008
© 2008 Organizing Committee of MIE 2008. All rights reserved.

Improving Patients Privacy with Pseudonymization

Thomas NEUBAUER and Bernhard RIEDL
Secure Business Austria, Vienna, Austria

Abstract. e-Health requires the sharing of patient related data when and where necessary. Electronic health records promise to improve communication between health care providers, thus leading to better quality of patients' treatment and reduced costs. As highly sensitive patient information provides a promising goal (e.g., for attackers), there is an increasing social and political pressure to guarantee patients privacy. This paper presents the new system PIPE (Pseudonymization of Information for Privacy in e-Health), that differs from existing approaches in its ability to securely integrate primary and secondary usage of health data.

Keywords. Privacy, e-health, Security, EPR-CPR-EMR, Smart Cards

Introduction

The discussion of privacy is one of the fundamental issues in health care today and a trade-off between (i) the patient's requirement for privacy as well as (ii) the society's need for improving efficiency and reducing costs of the health care system. Electronic health records (EHR) were introduced over the past several years as a method for improving communication between health care providers and access to data and documentation, leading to better clinical and service quality [10]. The EHR promises massive savings by digitizing diagnostic tests and images. A study by the non-profit research organization Rand Corporation found out that adopting the EHR could result in more than $81 billion in annual savings in the US, if 90% of the health care providers used it [5]. However, highly sensitive data is stored and handled in nationwide or european-wide medical sys-tems that are often available over the Internet and hardly protected. As a result of the high sensitivity of medical data and due to an almost endless list of security breaches revealing patients' data (cf. [2]), there is an increasing social, legal and political pressure to prevent the misuse of health data.

It is the patients right to demand privacy, because the disclosure of medical data may cause serious problems for the patient. Insurance companies or employers could use this information to deny health coverage or employment. Therefore, legal acts demand the protection of health data (HIPAA [22], Directive 95/46/EC [6], Article 29 Working Party [7], Article 8 [3] of the European Convention for the Protection of Human Rights and Fundamental Freedoms, domestic acts, e.g., the Austrian Data Protection Act [15]). This paper evaluates existing privacy enhancing techniques (PETs) used for securing medical data and presents a novel approach for granting privacy through encrypted pseudonyms and authorization via encryption named PIPE

(Pseudonymization of Information for Privacy in e-Health). This new system differs from existing approaches in its ability to securely integrate primary and secondary usage of health data. By showing new concepts for data sharing, authorization and data recovery in case the user loses her access key, it provides a solution to security shortcomings of existing approaches.

1. Background

The increasing fear of data abuse as well as the need for compliance to the legal demands lead to the development of a variety of techniques for protecting patients' identity and privacy. In order to protect patients' privacy when using and transferring medical records a variety of approaches is currently used (cf. Figure 1 for a comparison of existing approaches regarding secondary use):

- Anonymization (cf. [14,12,21]) is the removal of the identifier from the medical data. It is realized by deleting the patient's identification data and leaving the anamnesis data for secondary use. As this approach does not establish a link be-tween the anonymized data and its associated individual, it is the most secure way for granting privacy. Although this approach is often used in research projects due to its simplicity, it has the major drawback that primary use cannot directly profit from the results made in the research project (e.g., patients cannot be in-formed about actual findings such as newly developed medical treatment or major changes in the healing progress).
- Depersonalization (cf. [14]) is a technique similar to anonymization and comprises the removal of as much information as needed to conceal the patients' identity.
- Encryption (cf. [17]) assures patients' privacy by encrypting the anamnesis data with the patients' private key. As medical data tends to be very large (e.g., the image size of a x-ray is 6 MB, for a mammogram 24 MB or for a computer tomography scan counts up to hundreds of MB) encryption is a highly time-consuming operation. Beside that, encrypted data cannot be used for research projects (secondary use) without explicit allowance by the patient who has to decrypt the data and, thus, discloses his identity.
- Role-based access control (RBAC) (cf. [16]) can be used for restricting the data access to authorized persons. However, role-based access control models can be by-passed or compromised and persons with administrative roles still have un-restricted access to the whole database. Therefore, this technique does not pro-vide enough security for protecting sensitive health data, because attackers - no matter if working inside the system or getting access from outside the system -could get full access to the database including the relations between patient's identifiers and their medical data.
- Pseudonymization (cf. [11,20,4,8,13,12,19]) allows an association with a patient only under specified and controlled circumstances. It is a technique where iden-tification data is transformed into a and afterwards replaced by a specifier which cannot be associated with the identification data without knowing a certain secret. Existing approaches have shortcomings such as the concealment of the applied algorithms or the use of centralized patient-pseudonyms lists. Relying on a list is not secure, because an attacker, who gains access to this list, could establish an unauthorized relation between the

patient's identifier and his medical data. Existing approaches also neglect to provide efficient and secure fall-back mechanisms for recovering the key in order to re-establish access to the data if the patient lost his smart card and often depend on a single pseudonym.

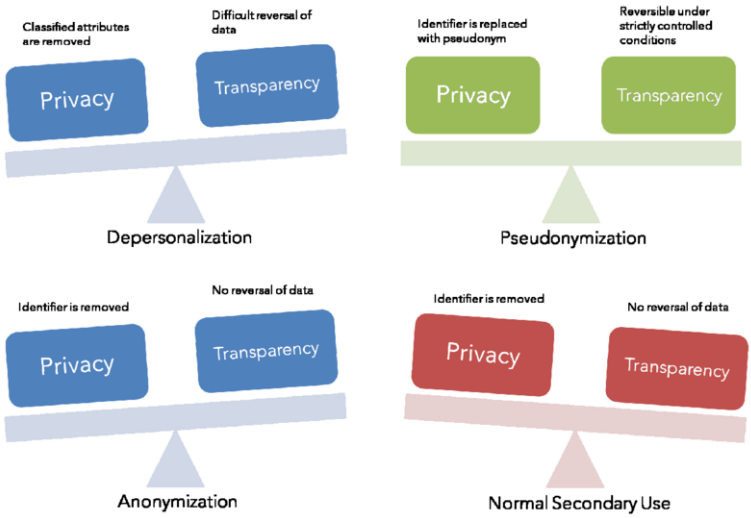

Figure 1. Comparison of existing approaches

2. Pseudonymization of Information for Privacy in e-Health

The goal of our architecture is to gain the optimal trade-off between security on the one hand and usability and performance on the other hand. PIPE comprises the following components: The Storage is divided into two separate storage systems (e.g., databases), where one is related to identification data and the other one is related to pseudonymized data as well as the associated pseudonym *PSN*. The central logic (e.g., a server) provides an interface between the central storage and the clients for the purpose of saving and loading the data. Note, that all private keys are identified as *e* (e.g., the patients' inner private key will be named *e'A*) and all public keys are denoted with *d*. The patient *(A)* has full access to his data via the central logic by using his security token (e.g., smartcard with a PIN). The patient may provide relatives *(B)* with his inner private key, which will then be encrypted with the relative's symmetric key. By doing this, the relative gets access to all data of the patient, until the patients' inner private key is changed. A health care provider *(C)* can be authorized by the patient to access a subset of his anamnesis datasets. The health care provider may also share one or more entries in the pseudonymized database with the patient. The research lab has only access to the concealed data *CD* on the server system via the central logic for the purpose of analysis needed for improving the efficiency of clinical trails, the medical treatment, or medication. The operator(-team) *(O)* holds secrets on behalf of the system. This role assures that if a patient loses or destroys his smartcard, the access to the system can be restored by a team of operators.

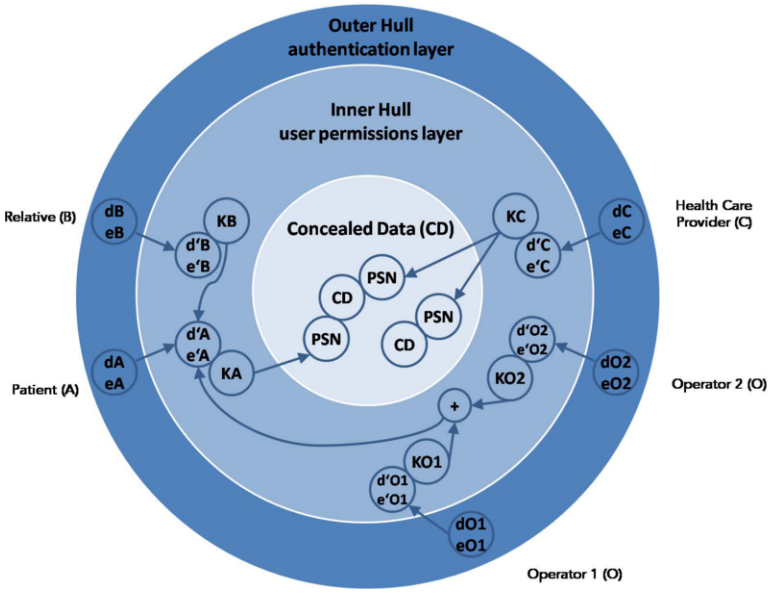

Figure 2. The Hull Architecture of PIPE

Figure 2 shows the hull architecture of PIPE. In the most outer layer (user permissions layer) every user possesses a security token (e.g., in our prototype we used smart cards as security tokens) to access the secrets of the next inner hull. The anamnesis data is stored pseudonymized in the most inner hull (concealed data layer).

Any medical dataset is associated with one or more unique pseudonyms. As the patient is the owner of the data, she is the only person who holds the so-called root-pseudonym PSN_0. All other pseudonyms PSN_j are disjunct for any patient, health-care-provider and anamnesis combination. If, for example, two health care providers have been authorized to access a specific anamnesis *CD*, three pseudonyms (PSN_0, PSN_1 and PSN_2) exist. All of these pseudonyms are stored encrypted with the particular users' inner symmetric keys, whereas the plain-text medical data is associated with the plain-text pseudonyms. Thus, the pseudonymization can be reversed by using the patient's inner symmetric K_A or any authorized health care provider's *C* inner symmetric key K_C. To get access to these particular keys, the authorized users' inner private key has to be used. If a patient *A* shares her secret of the inner hull, she consequently provides access to all her data, if not additionally revised by an access control model. We define two main roles which may hold an encrypted copy of the patient's inner hull secret, her inner private key *e'A*. Firstly, a relative *B* may encrypt the patient's inner private key with her inner public key *d'B*. Thus, she is also able to decrypt the patient's inner symmetric key, until the patient would change it.

Secondly, as a user's smart card may be lost, destroyed, stolen, compromised or just worn-out, there is the need to keep a backup of the user's inner private key, because otherwise the user's data would not be accessible any more. This backup keystore has to be secured and protected against fraud. In our prototype we applied Shamir's threshold scheme for securely sharing secrets [18] between a set of operators *O*, who are randomly assigned to hold a part of the users' secrets. Since users need a fall-back mechanism in case they have lost their smart card, operators *O* hold the users'

inner private keys on behalf of the patient. To avoid misuse of these rights, we applied a two-step variant of Shamir's threshold scheme [18]. Following Shamir, two parameters can be defined for sharing a secret, (i) the number of shares *n* and (ii) the amount of shares *k*, that are necessary to re-establish the certain secret. The higher the number of issued shares compared to the number of shares in total, the higher the security, assuming the operators are randomly assigned, each holding one share. Of course, decrypting operations conducted by humans cause higher costs than performed by machines. To decrease the costs for establishing a backup keystore, we propose a combination of human operators and machine operators, for example Hardware Security Modules. Secure authorizing is another vital function of our system to assure that the users are in full control of their actions at any time. If there is for example the need to authorize an additional user for an anamnesis, another pseudonym has to be linked with the certain anamnesis, which the data owner and the additional user hold together.

3. Conclusions

Health care providers require the sharing of patient related data (e.g., using EHRs) in order to provide efficient patients treatment. The implementation of EHRs does not only promises a higher level of service quality for the patients, but also reduces costs for social insurance systems and therefore for the society. As highly sensitive and personal infor-mation is stored and shared within these highly interconnected systems, there is increas-ing political, legal and social pressure to guarantee patients' privacy. This paper gave an overview of the shortcomings of existing approaches, such as their dependence on a cen-tralized patient-pseudonyms list, a life-long pseudonym or the concealment of the used algorithm. Based on these shortcomings this paper presented a secure architecture for the combined primary and secondary usage of health-related data. Relying on the encryption of the systems relations, PIPE assures that the patients are in full control of their data. The approach provides research organization with data for improving medical treatment or clinical paths. The proposed fall-back mechanism allows recovering the patient's secret key in case he lost his smart card. Workflows that demand the authorization of one of the systems' participants demand a the use of a secure viewer (cf. [9,1]) in order to guar-antee confidentiality, integrity and non-repudiation (e.g., that only the intended person is authorized). The information about the person that should be authorized must be pre-sented to the user without any manipulations (e.g., committed by a man-in-the-middle). Therefore, further research will extend our concept by integrating a secure viewer mech-anism and present the results of some case studies. The case studies will focus on the integration of our concept into existing workflows and systems in the health sector in compliance with legal requirements.

Acknowledgments

This work was performed at Secure Business Austria, a competence center that is funded by the Austrian Federal Ministry of Economics and Labor (BMWA) as well as by the provincial government of Vienna.

References

[1] Adil Alsaid and Chris J. Mitchell. Dynamic content attacks on digital signatures. *Information Management & Computer Security*, 13(4):328–336, 2005.

[2] Privacy Rights Clearinghouse. A chronology of data breaches.

[3] Council of Europe. *European Convention on Human Rights*. Martinus Nijhoff Publishers, 1987.

[4] G.J. de Moor, B. Claerhout, and F. de Meyer. Privacy enhancing technologies: the key to secure commu-nication and management of clinical and genomic data. *Methods of information in medicine*, 42:148– 153, 2003.

[5] Frank R. Ernst and Amy J. Grizzle. Drug-related morbidity and mortality: Updating the cost-of-illness model. Technical report, University of Arizona, 2001.

[6] European Union. 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 of the European Communities*, L 281:31–50, 1995.

[7] European Union, Article 29 Working Party. Working document on the processing of personal data relating to health in electronic health records (EHR), February 2007.

[8] J.R. Gulcher, K. Kristjansson, H. Gudbjartsson, K., and Stefanson. Protection of privacy by third-party encryption in genetic research. *European journal of human genetics*, 8:739–742, 2000.

[9] Hanno Langweg. Malware attacks on electronic signatures revisited. In J. Dittmann, *Konferenzband der 3. Jahrestagung Fachbereich Sicherheit der Gesellschaft für Informatik*, pp. 244–255, 2006.

[10] S. Maerkle, K. Koechy, R. Tschirley, and H. U. Lemke. The PREPaRe system – Patient Oriented Access to the Personal Electronic Medical Record. In *Proceedings of Computer Assisted Radiology and Surgery, Netherlands*, pages 849–854, 2001.

[11] Robert L. Peterson. Patent: Encryption system for allowing immediate universal access to medical records while maintaining complete patient control over privacy. *US Patent US 2003/0074564 A1*, 2003.

[12] Andreas Pfitzmann and Marit Koehntopp. Anonymity, Unlinkability, Unobservability, Pseudonymity, and Identity Management -A Consolidated Proposal for Terminology. In *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2005.

[13] Klaus Pommerening and Michael Reng. *Medical And Care Compunetics 1*, chapter Secondary use of the Electronic Health Record via pseudonymisation, pages 441–446. IOS Press, 2004.

[14] Alan Rector, Jeremy Rogers, Adel Taweel, David Ingram, Dipak Kalra, Jo Milan, Peter Singleton, Robert Gaizauskas, Mark Hepple, Donia Scott, and Richard Power. Clef -joining up healthcare with clinical and post-genomic research. In *Proceedings of UK e-Science All Hands Meeting*, pages 203–211, 2003.

[15] Republic of Austria. Datenschutzgesetz 2000 (DSG 2000), BGBl. I Nr. 165/1999, 1999.

[16] R.S. Sandhu, E.J. Coyne, and C.E. Feinstein, H.L.and Youman. Role-based access control models. *Computer*, 29(2):38–47, 1996.

[17] Volker Schmidt, Werner Striebel, Heinz Prihoda, Michael Becker, and Gregor De Lijzer. Patent: Verfahren zum be-oder verarbeiten von daten. *German Patent, DE 199 25 910 A1*, 2001.

[18] Adi Shamir. How to share a secret. *Commun. ACM*, 22(11):612–613, 1979.

[19] Ka Taipale. Technology, Security and Privacy: The Fear of Frankenstein, the Mythology of Privacy and the Lessons of King Ludd. *International Journal of Communications Law & Policy*, 9, 2004.

[20] Christian Thielscher, Martin Gottfried, Simon Umbreit, Frank Boegner, Jochen Haack, and Nikolai Schroeders. Patent: Data processing system for patient data. *Int. Patent, WO 03/034294 A2*, 2005.

[21] Denise Thomson, Lana Bzdel, Karen Golden-Biddle, Trish Reay, and Carole A. Estabrooks. Central Questions of Anonymization: A Case Study of Secondary Use of Qualitative Data. *Forum Qualitative Social Research*, 6:29, 2005.

[22] United States Department of Health & Human Service. Hipaa administrative simplification: Enforcement; final rule. *Federal Register / Rules and Regulations*, Vol. 71, No. 32, 2006.