Protecting Patient Privacy when Sharing Medical Data

Stefan Benzschawel
CR SANTEC - Public Research Center Henri Tudor
Luxembourg-Kirchberg, Luxembourg
stefan.benzschawel@tudor.lu

Marcos Da Silveira
CR SANTEC - Public Research Center Henri Tudor
Luxembourg-Kirchberg, Luxembourg
marcos.dasilveira@tudor.lu

***Abstract*— This paper describes a national eHealth platform concept with a multi-level privacy protection in order to improve the security and privacy of medical information on their storage locations as well as during the exchanging/sharing processes. The key idea is to classify and split-up data into different servers. A Trusted Third Party server manages personal identifying data together with the related pseudonyms while the medical information server manages the related medical data assigned to pseudonyms. The well known IHE-XDS profiles are enriched by Public Key Infrastructure, symmetric and asymmetric encryption together with pseudonymization methods. IHE-XDS promote the interoperability level and the extensions increase the security level.**

Keywords*— *eHealth; Patient Privacy; Electronic Health Records; Secure Patient Data Storage

I. INTRODUCTION

Healthcare technologies are moving from isolated and autonomous solutions to more interoperable ones. The main expectations of this change are to provide better ways to exchange and share medical information and to improve the quality of services offered to the patients.

In this context, medical data is supposed to be available online where healthcare professionals can access it at any time and from any place. Basically, it will be transmitted over Internet, dedicated Virtual Private Networks (VPN), and hospital networks. The on-line access to medical information can have two major consequences: it can support healthcare professional to take better decisions; it can increase the risk of loss of privacy and malicious attacks. The goal of designing and implementing eHealth platforms is to reinforce the former consequence and to reduce or eliminate the second one. This paper focuses on the strategy to widely reduce the malicious attacks' risk and to assure the privacy of patients during the storing and exchange (sharing) of medical information by using the eHealth platform.

Some cryptographic protocols have proved their efficiency to provide data-security for communications over networks but they do not fully prevent attacks to users' computers or servers. An eHealth platform has to deal with these risks, control authentication, authorization, and integrity. Several countries are implementing different solutions to satisfy these needs, but the evolution of the applications, methods and laws had forced some of them to review partially or completely their approaches.

The terms "central" and "decentral" mostly refer to the location of the information repositories. In enlarging this interpretation towards different components of a system, the term "central" refers to a system where the components are in one location, managed by one staff of administrators. The term "decentral" then refers to a distributed system. The security advantage of decentral systems is that an attacker will get only a part of the stored information. The disadvantage is that the components (satellites) of the distributed system may not be protected in the same "best" way, as one can do for a centralized system.

The proposed solution respects both aspects – (1) avoid a single attack point and (2) data-security for the satellites' data. If the information stored in one satellite is unusable for an attacker as long as information from other satellites is missing, thus the hacked information of one satellite is worthless alone. This paper describes a secure IT-platform based on this idea. It protects the stored information against external intruder attacks as well as against internal administrator attacks. The layout is based on the IHE-XDS [1] profile, extended with pseudonymization, encryption, and signature functionalities.

Patients' data are distributed within the system in two main parts: One for storing the medical information under pseudonyms and the other for mapping the pseudonyms to the patients' identity data. The benefit is: if one part of data is stolen, the information is useless. Neither the mapping table nor the medical database with pseudonyms is really useful alone.

In the case where medical data contains additional person identifying information, like a name printed on an X-ray picture, then the medical data (i.e., the X-ray) is encrypted and stored under a pseudonym.

The access to the stored information is realized with a web application. As the web-server in the Internet is a high security risk, the patient's identifying data are hidden against the web-server. Illegal server logs on the web-server are useless. Also to avoid illegal web-server logs, the transferred medical results are encrypted on their way over the web-server.

The platform is protected against unauthorized access by multiple security levels. The initial login is done with a

 ISBN: 978-1-61208-119-9

personal ID-card. A user&role directory guards the legal access to the system.

Finally, an elaborated consent management protects any undesired access on the base of the patients' will. The special case of an information access during an emergency situation of the patient implies sending an information about the data access to the patient, his family doctor or any other named contact person.

In the next section, some related works are shortly presented and discussed. Section III introduces the architectural approach.

II. Related Work

Data stored in and transmitted through an Internet-based platform are confronted with a set of attack possibilities. In health care domain, medical data and personal data can be exchanged and managed by services provided in a platform. It also includes privacy and data protection. Patients want to be sure that their personal and medical information is not misused. They want to know how their data is utilized, disclosed, and protected, and the degree of control they will have over the dissemination of this information. They are also worried about possible undesirable economic and social consequences from the misuse of such information [2][3]. Users of healthcare services are unwilling to have their personal information distributed other than for purposes of clinical care and they would like to be consulted before their information is released. The right to decide which personal information can be communicated to others and under which conditions constitutes their privacy rights and need to be implemented in the platform system. Assuring privacy implies that the platform needs to deal with, at least two attack risks eavesdropper and server intruders or curious insiders [4]. Three potential types of attackers are described in [4]: *Client intruder*, which attack the client computer (e.g., trojan); *Eavesdropper*, which compromise or owns a subset of communication's nodes to collect and analyze messages that are routed over them; *Curious insider or server intruder*, this attackers have administrative privileges and can access all data in the server.

For the client intruders' risk, we assume that users are responsible for the protection of their own system and of data saved in their computer. The access to the platform is protected by a system based on the electronic cards that provides identification and signature services. If an intruder steals the identity of the user he will need to have his card and know his password to use the platform. Other countries have also adopted electronic cards for health data management and patient identification (eCard in Austria[5], eGK in Germany[6], Vitale in France[7], etc), this type of card has shown its efficacy in banking domain and are widely accepted by users.

The other two types of attacks are directly related to the security strategy of the platform. Cryptography (e.g., Public Key Infrastructure) is commonly used in eHealth platforms to avoid eavesdropper, however a communication protocol should be defined to avoid that encrypted data and decryption keys cross the same node without a specific protection. The Belgium platform deals with this problem by implementing an end-to-end communication [8], then the private key is expected to never leaves the client computer. However, this solution does not allow sharing data when the receiver is unknown. For example, a prescription cannot be accessed by a pharmacy if the pharmacy was not chose by patient/doctor at the moment of the e-prescription creation.

In Luxembourg [9], the eHealth platform has been designed to store (temporarily) medical data, and users will be able to access this data. In this case, the information cannot be encrypted with the public key of the (unknown) receiver and saving unencrypted data will open a door for server intruders or curious insiders attacks. The protocol proposed in this paper deals with this situation using symmetric encryption associated to asymmetric encryption for the symmetric keys [10], an identity and role control system and a pseudonymization of unencrypted data [11]. This encryption technologies are well known by Network administrators, however, associating it with pseudonymization techniques are not usual, as much as we know the proposed solutions use proprietary message structures. As semantic interoperability is an important step to promote sharing/exchange of information in the medical domain, the contribution of our work is the association of these technologies within the IHE-XDS profile.

Ideally, privacy is assured if a consumer uses a resource or service without disclosing his consumer identity; the resource or service can be used multiple times without others being able to link these uses together (unlinkability) or observe that this resource is being used (unobservability) [2]. In medical domain, it can be illustrated by a doctor accessing a set of data of one patient stored in several repositories. The system need to assure that nobody else can know that all these data belongs to the same patient (unlinkability). Or, when a patient access his own data, the system should assure that nobody will observe it (unobservabiity), because if the user (patient) is associated to the data, the unlinkability criteria is lost. The encryption and the pseudonymization techniques can not solve this problem. An organizational strategy is necessary to improve the privacy of patients, and this is another contribution of the paper.

The eHealth architecture detailed in the next section shows how privacy, hiding users' identity and assuring authenticity/integrity of documents and messages can be established. The approach is based on a multi-level architecture where: users are authenticated by a trust authority and associated to a set of rights access; data are pseudonymized; non-anonymized documents are encrypted; strict sequences of activities are provided; and messages are stored encrypted for audits purposes.

 ISBN: 978-1-61208-119-9

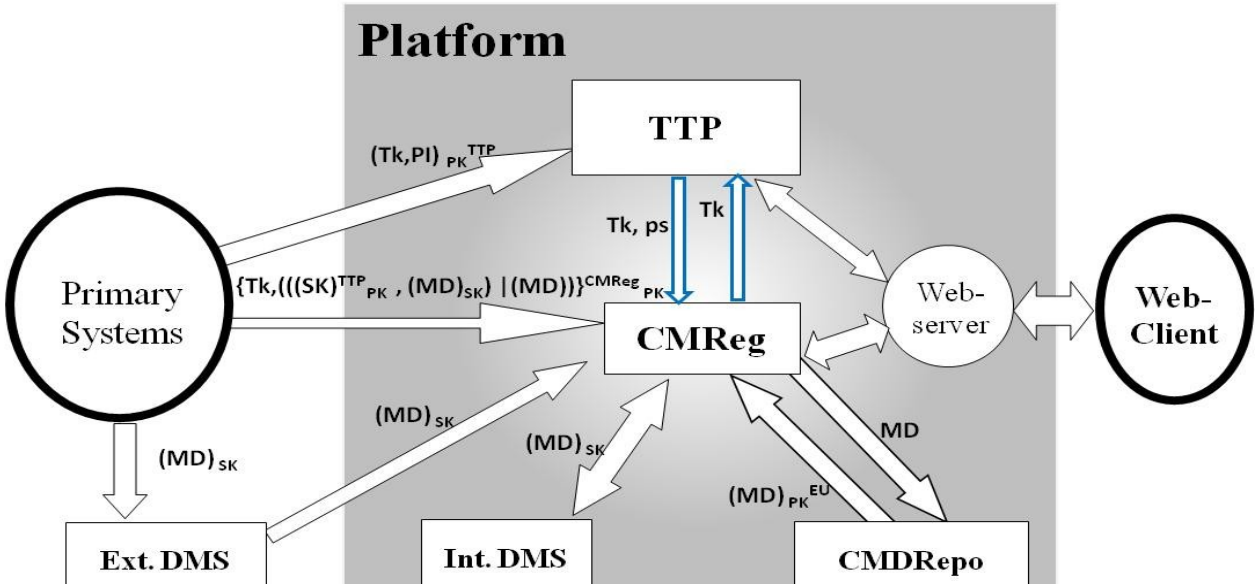

Figure 1: Architecture of the Platform

III. Architectural Overview

The main components of the proposed eHealth platform are presented in Figure 1. Some components (e.g., LDAP, CA, STS, etc.) are omitted to improve the visibility and the architecture explanation. The technology used to implement these components are out of the scope of this paper. The architecture was defined to support centralized and decentralized information repositories, based on the IHE XDS infrastructure profile with a central registry, one or more centralized repositories and one or multiple decentralized repositories.

The heart of this platform is the *Central Medical Registry* (CMReg). Each document provided by the primary systems is registered in the CMReg with its physical location in one of the repositories. For illustration, Figure 1 shows two centralized repositories: (1) the *Centralized Medical Data Repository* (CMDRepo) stores unencrypted information without any person identifying data; and (2) the *Internal Document Management System* (Int. DMS) that stores encrypted medical information, which may contain, as well, person identifying data. External storage of data is also supported by the platform. For example, Primary Systems may decide to use their own repositories (Ext. DMS), placed in a DMZ (Demilitarized Zone) from their network. Those external repositories always contain encrypted information.

This organization allows normalizing data storage and data retrieving within all data repositories. CMReg keeps meta-data of all information registered in the platform, what makes the CMReg a potential target for malicious attacks. Protecting the CMReg is one priority of our security strategy. A set of components is combined to improve the security of the system. In order to describe the data exchange protocol, it is assumed that an existing Public Key Infrastructure (PKI) is in place and each registered entity has a public/private identity key pair. The notations introduced in Table 1 are used.

Table 1: Notations

Notation	Meaning
PI	Patient Identifying Data (name, address, sex, …)
MD	Medical Data (lab results, diagnosis, …)
EU	End User. Can be patients, health providers, researchers, etc.
SK	Symmetric key
Tk	Token
ps	Pseudonym
$(m)_{SK}$	Message encrypted with a symmetric key
$(m)^{U}_{PK}$	Message encrypted with the public key of a user *U*

A. User identification

Two groups of users are considered for the eHealth platform, according to the role that they play: (1) Primary Systems, who uses the platform to send medical data produced by healthcare providers (ex., laboratory results, x-rays, or discharge letters) via the "Push" web-service; (2) End Users, healthcare professionals or patients that use the platform to acquire stored medical information.

The procedure to use the platform is the same for both groups. Users need an electronic card (eID for short) for authentication and for data integrity (through e-signature of documents).

The access to the system requests the following steps:

1. Users holding their eID to request an "entry ticket" to the Secure Token Service (STS). The user can request an entry token via Web (i.e., as data consumer through a Web-client – right side of Figure 1) or via an Intranet (i.e., as data producer through HealthNet – left side of Figure 1). The request message is signed on user's side and encrypted with the STS public key.

$$\text{User} \rightarrow \text{STS: (eID)}^{\text{STS}_{\text{PK}}}$$

2. STS verifies the signature with the certification authority (CA). If refused, the user's access is denied.

$$\text{STS} \rightarrow \text{CA: (eID)}^{\text{CA}_{\text{PK}}}$$
$$\text{CA} \rightarrow \text{STS: (Check result)}^{\text{STS}_{\text{PK}}}$$

3. STS requests access rights information to Lightweight Directory Access Protocol (LDAP) manager. The answer is a set of roles that this user can play.

$$\text{STS} \rightarrow \text{LDAP: (eID)}^{\text{LDAP}_{\text{PK}}}$$
$$\text{LDAP} \rightarrow \text{STS: (roles)}^{\text{STS}_{\text{PK}}}$$

4. STS prepares the entry ticket, encrypts it with the user's public key and sends it to the user.

$$\text{STS} \rightarrow \text{User: (entry ticket)}^{\text{User}_{\text{PK}}}$$

For the client, the entry ticket will give the access to a set of Web-applications in the Web-Server and for the Primary Systems, it will allow them to use the "Push" web-services provided by CMReg system.

This protocol requires that users are pre-registered at STS, that they have an eID recognized by a certification authority (CA) and that he uses this eID during the whole process (entry and service request). Data (e.g., the entry ticket) will be rendered encrypted and the user needs his private key to decrypt. This strategy protects users from eID stealers.

B. Pseudonymization

As medical data are registered in the CMReg, they are associated to pseudonyms and stored in one of the data repositories. The mapping between pseudonyms and the corresponding person identifying data is stored in a Trusted Third Party (TTP). We use the term TTP for the mapping software and TTP driver for the organization that operates this software. The mapping between the person identifying data and the pseudonyms must never be disclosed. To assure this, a «token» is used and all communication is encrypted. The psedonymization service can be summarized in the following 6 steps:

1. The Primary Systems (PrS) provides a clean separation of person identifying data and its related medical data (i.e., two separated documents are created). The medical data may be unencrypted but without any person-identifying information, or encrypted.
2. The person identifying data is sent to the TTP together with a token. $\text{PrS} \rightarrow \text{TTP: (Tk,PI)}^{\text{TTP}_{\text{PK}}}$
3. The medical data (or a reference to the medical data) is sent to the CMReg (Push service) with the same token. The document itself is stored in one of the repositories. $\text{PrS} \rightarrow \text{CMReg: (Tk, MD)}^{\text{CMReg}_{\text{PK}}}$
4. The TTP generates a pseudonym, stores it besides the person identifying data and waits for the CMReg asking for the pseudonym.
5. The CMReg sends a request to TTP with the "token" and gets back the generated pseudonym:

$$\text{CMReg} \rightarrow \text{TTP: (Tk)}^{\text{TTP}_{\text{PK}}}$$
$$\text{TTP} \rightarrow \text{CMReg: (Tk,ps)}^{\text{CMReg}_{\text{PK}}}$$

6. CMReg establish a mapping between the pseudonym and the (encrypted) document with the medical data. The pseudonym becomes part of the meta-data of the document and is not visible outside of the platform;

Additional security packs can be used to improve the privacy of patients. For example:

- Scheduled pseudonym exchange: The pseudonyms will be exchanged on a regular basis each hour or if necessary in shorter intervals. The stolen mapping table of a hypothetical evil TTP administrator only works if the hypothetical evil PMIP administrator has stolen the medical databases during the same time interval. If this extension gets necessary further elaboration concerning the switching time has to be done.
- Multiple pseudonymization: To further enlarge the trust level, multiple pseudonymization steps are possible. The first pseudonymization service maps real identities to pseudonyms. The second pseudonymization service maps the first pseudonym to a second pseudonym. The n-th pseudonymization service maps the (n-1)-th pseudonym to an n-th pseudonym. Each pseudonymization mapping is hosted by an independent trusted N-th party.
- A combination of scheduled pseudonym exchange and multiple pseudonymization with different pseudonym exchange intervals of the different levels is possible.

 ISBN: 978-1-61208-119-9

C. Encryption/Re-Encryption

When the separation of person identifying data and the related medical data is not possible (e.g., X-ray image), the privacy is guaranteed by a combined encryption strategy. It consists of 5 steps:

1. The medical document (MD) is symmetrically encrypted with a symmetric key generated by the PrS – one for each document, respectively;
$$\{(MD)_{SK}\}$$
2. The symmetric key is encrypted with the public key of TTP;
$$\{(SK)^{TTP}{}_{PK}\}$$
3. The encrypted document and the encrypted key is stored together in one of the repositories.
4. When requested by an authorized user, the encrypted key of the document is separated from the document, sent to TTP, which will be in charge of re-encrypting the key with the public key of the legal requester, and regrouped with the document:
$$CMReg \rightarrow TTP: ((SK)^{TTP}{}_{PK}, EU)^{TTP}{}_{PK}$$
$$TTP \rightarrow CMReg: ((SK)^{EU}{}_{PK})$$
5. Both, the encrypted document and the re-encrypted key, are sent to the end-user.
$$CMReg \rightarrow EU: \{(SK)^{EU}{}_{PK}, (MD)_{SK}\}$$

This distributed encryption/re-encryption strategy prevents both insiders' server attacks and eavesdroppers. The repositories store the encrypted documents with the encrypted keys. The re-encryption of the encrypted symmetric keys is done at the TTP side. The TTP never has access to the encrypted document while the repository never has access to a disclosed symmetric key. For eavesdroppers of the repository or eavesdroppers of the TTP the same argument holds like for the corresponding administrators. Only with simultaneous access to TTP and repository the information can be disclosed. In this process, the pseudonym of the patient can differs from one primary source to another, what improve the unlinkability of the solution.

D. Hiding information from the servers' administrators

Two types of files with medical data are stored in the platform one unencrypted/pseudonymized and the other encrypted/pseudonymized. At least 4 servers compose the platform infrastructure (TTP, CMReg, Repositories, Web-server). TPP and Repositories are protected by the trick described above. The Web-server is often the main target of attacks because it is used to transmit data to/from the end user. A malicious administrator may install an illegal logging, catching the requests containing patient names and catching the results containing medical data for those patients. With a simple strategy, after cumulating this log information, the administrator may associate the set of health data with patients' identity. To prevent this, patient identifying data are encrypted with the public key of the TTP. Then the web-server only transmits the information to the TTP. And the TTP has one additional step to do: it has to decrypt the patient identifying data. Then it continues by looking-up and providing the pseudonyms and waiting for the CMReg request for the pseudonym-list. An analogous tunneling method is applied for result transmission over the web-server to the receiver.

E. Consent and User Management

A secure token service with a healthcare related LDAP guards the access to the whole system. Users need to be pre-registered and associate to a set of access rights to use the applications of the system. An elaborated consent management system protects medical information from any unwished access on the basis of the patient's will. For example:

- for all documents, for an episode, for a medical case;
- for all doctors, for all doctors of a special discipline, for named doctors;
- for exchange over borders;
- for access in emergency case;
- …

A specific consent description language [12] has been proposed to declare consents. CMReg checks the conformance of an access first with the patients' consent declaration for each requested document. Patients can access the system via web-applications and their identity will be substituted by a pseudonym define by TTP (following the same process described in 3.2).

F. Trustful statistics

Pseudonymized results in the platform offers the possibility of using data for statistics analysis without the risk of data protection violations. Therefore a preparation process is necessary to exchange the internal used pseudonyms by other pseudonyms created for the statistics purposes. Internal used pseudonyms are supposed to be hidden from external users. Statistics analysis can use a predefined set of not encrypted medical data that must not contain patient identity information. If further statistical research has to be done, where personal data like age and sex are necessary, exceptions can be created. But, it may require special authorizations from public authorities in order to guarantee citizens' privacy.

 ISBN: 978-1-61208-119-9

IV. Conclusion and Future Work

This paper has presented an architecture for eHealth platforms that combine different methods for data protection in order to improve the security level and assure the privacy of patient's data.

The architecture's concept was developed based on standards protocols and proposes some extensions for multi-level privacy protection. The extensions consist mainly on the communication with the Trusted Third Party server and are shown to be necessary when considering potential intern attacks (malicious administrators). A "ticket" based protocol is proposed to assure authentication of users. It is associated with an electronic card (provided by a certification authority) that offers signature and identification services.

The architecture was designed to promote exchange and sharing of medical data and to collect/store data for statistics finalities. Thus, the collected data could not be encrypted, but the identity of the patients is never exposed. The proposed approach uses pseudonymization methods for hiding patient's identifying data. But, during the data exchange process, even pseudonymized data are encrypted using PKI solutions to avoid eavesdroppers attacks. As the platform was not conceived to provide P2P communication, a strategy to safely store unanonymized data was necessary. An association of symmetric and asymmetric encryption is proposed, which involves at least two different servers to provide the necessary information to the end user. This approach has shown to be efficient against insiders' servers attacks and against client intruders that try to steal the client identity.

Future works are planned to improve the identity protection when the patient should be able to access his own data. This particular increase the risk of eavesdroppers attacks over the web-server because the identity of the patient is known (equal to the requester identity). We are also working on the implementation of the platform and on the validation with case studies specified with user groups.

V. Acknowledgements

The authors thank their partners from the Health Ministry of Luxembourg for their very helpful advises and for providing insights into organizational and legal aspects of the eHealth platform.

VI. References

[1] IHE Integrating the Healthcare Enterprise. IT infrastructure technical framework vol.1 (iti tf-1) integration profiles. Technical report, 2007. Last access: 10/12/2010 http://www.ihe.net/Technical_Framework/upload/IHE_ITI_TF_4_0_Vol1_FT_2007_08_22.pdf

[2] G. Bansal, F. Zahedi, and D. Gefen, "The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online," Decision Support Systems, vol. 49, 2010, pp. 138-150.

[3] R. Au and P. Croll, "Consumer-Centric and Privacy-Preserving Identity Management for Distributed E-Health Systems," Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008), 2008, pp. 234-234.

[4] D. Slamanig and C. Stingl, "The Degree of Privacy in Web-based Electronic Health Records," ECIFMBE 2008, J. Vander Sloten, P. Verdonck, M. Nyssen, and J. Haueisen, (Eds.), 2008, pp. 974-977.

[5] Last access: 10/12/2010 http://www.chipkarte.at/portal27/portal/ecardportal/start/startWindow?action=2&p_menuid=51682&p_tabid=1

[6] Last access: 10/12/2010 http://www.telematik-modellregionen.de/content/

[7] Last access: 10/12/2010 http://www.sesam-vitale.fr/programme/programme_eng.asp

[8] Last access: 10/12/2010 https://www.ehealth.fgov.be/binaries/website/en/20100531_en.ppt

[9] S. Benzschawel, H. Zimmermann, M. Da Silveira, U. Roth, A. Jahnen, "IT infrastructure for National Electronic Health Records in Luxembourg – Acceptance occurs when benefits outweigh disadvantages." Global Telemedicine and eHealth Updates: Knowledge Resources, vol. 3, Malina Jordanova and Frank Lievens (Eds.), 2010, pp. 141-145

[10] D. Galindo and E. R. Verheul, "Pseudonymized Data Sharing". Privacy and Anonymity in Information Management Systems. J. Nin, J. Herranz (Eds.). Series: Advanced Information and Knowledge Processing, vol. 0, Part 3, 2010, pp. 157-179, DOI: 10.1007/978-1-84996-238-4_8

[11] T. Neubauer and J. Heurix, "A methodology for the pseudnymization of medical data". In press. International journal of medical informatics, 2010, doi:10.1016/j.ijmedinf.2010.10.016

[12] C. Pruski, "e-CRL: A Rule-based Language for Expressing Patient Electronic Consent," eTelemed 2010, Second International Conference on eHealth, Telemedicine, and Social Medicine, 2010, pp.141-146.

 ISBN: 978-1-61208-119-9