A HYBRID APPROACH INTEGRATING ENCRYPTION AND PSEUDONYMIZATION FOR PROTECTING ELECTRONIC HEALTH RECORDS

Johannes Heurix
SBA Research
jheurix@sba-research.org
Austria

Michael Karlinger, Michael Schrefl
Johannes Kepler University Linz
{karlinger, schrefl}@dke.uni-linz.ac.at
Austria

Thomas Neubauer
Vienna University of Technology
neubauer@ifs.tuwien.ac.at
Austria

ABSTRACT
Federated Health Information Systems (FHIS) integrate autonomous information systems of participating health care providers to facilitate the exchange of Electronic Health Records (EHR), which improve the quality and efficiency of patients' care. However, the main problem with collecting and maintaining the sensitive data in electronic form is the issue of preserving data confidentiality and patients' privacy. Although multiple technical measures to restrict access to only authorized persons are implemented, they are usually aimed against external attackers. In this work, we propose to integrate pseudonymization and encryption to a hybrid approach which not only protects against external attackers, but also ensures that even potential internal attackers with full data access, like administrators, cannot gain any useful information.

KEY WORDS
Health Care Information Systems, Medical Data Storage and Compression Techniques, Privacy, Federated Health Information Systems

1 Introduction

Have you ever had to do tedious examinations twice because you could not find the results from the previous examination or have you ever thought about why you are the one who has to take care of examinations and carry them from one doctor to the other? Most people who have to consult a general practitioner on a regular basis are faced with organizational inefficiencies of the health system. However, with the rise of information and communication technology and its application to the health care sector (often referred to as e-health), governments all over the world are on their way to reduce existent deficiencies. Electronic Health Records (EHR), for example, have the potential to improve communication between health care providers and access to data and documentation, leading to better clinical and service quality, and thus massive savings by digitizing diagnostic tests and images (cf. [3]). E-health and especially the EHR as one of its main pillars can revolutionize health care, but come at a price: privacy. With interconnected systems comes highly sensitive and personal information that is often available over the Internet and – what is more concerning – inadequately protected. Highly sensitive patient information provides a promising goal for attackers and is frequently demanded by insurance companies and employers. The disclosure of sensitive data, such as a history of substance abuse or HIV infection, could result in discrimination or harassment. In this discussion, privacy is often not the main concern, but surveillance and the effects it has - both positive and negative - on human values, relationships, and daily practice. Of course, a variety of legal acts demand the protection of health data. Historically, the definition of an individual's privacy as the "right to be let alone", was defined by the US Supreme Court in 1834. In 2006, the United States Department of Health & Human Services issued the Health Insurance Portability and Accountability Act (HIPAA) which demands the protection of patients' data that is shared from its original source of collection. In the EU, the processing and movement of personal data is legally regulated with Directive 95/46/EC. A citizen's right to privacy is also recognized in Article 8 of the European Convention for the Protection of Human Rights and Fundamental Freedoms. Additionally, domestic acts in many EU member states contain strict regulations concerning the processing of personal data. In order to protect patients' privacy when using, transferring, and storing medical records, a variety of privacy enhancing technologies (cf. [4] for a definition) have been proposed, but existing approaches often (i) do not comply with the legal demands and (ii) do not fulfill basic security requirements (cf. [17, 2]). Therefore, we propose to store the sensitive data in such a way that any potential internal or external attacker cannot gain any useful information, even if acquired full data access. In particular, we focus on pseudonymization and encryption. Because both techniques have their limitations, we propose to combine them to a hybrid approach for the protection of EHRs providing a priori data protection against data leakage.

2 Architecture of Federated Health Information Systems

This article covers the realization of the EHR within a Federated Health Information System (FHIS) that integrates

autonomous information systems of the participating health care providers. Figure 1 depicts the overall architecture of a FHIS proposed by the initiative for *Integrating the Healthcare Enterprise* (IHE). The IHE is a worldwide initiative by healthcare professionals and industry to improve interoperability of health information systems [1].

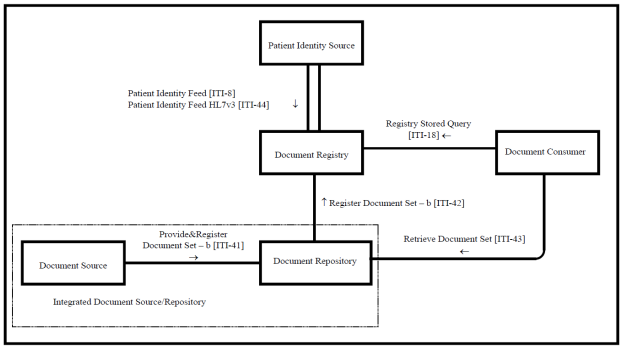

Figure 1. Cross-Enterprise Document Sharing – b [1]

The IHE Cross-Enterprise Document Sharing – b (XDS.b) profile consists of a centralized *patient identity source* and *document registry*, and decentralized *document repositories* where the actual documents are stored and assigned a unique record identifier (RID). To enable common processing, documents are persisted in agreed formats such as HL7 CDA and DICOM, and data is exchanged conforming to IHE IT Infrastructure (ITI) profiles and transactions. The patient identity source contains personal data about patients and health care providers (HCPs) and assigns a globally unique identifier to each patient (PID) and health care provider (HCP-ID). The document registry establishes the connection between a document in a document repository and the patient who owns the record.

To enable search, the document source supplies metadata to the EHR document. For example, the metadata of an EHR document includes the name of the HCP who created the record, the creation date and time, or the type of clinical activity. Subsequently the EHR document is stored in the document repository and registered in the document registry using the PID, the RID generated for the document, the location of the document repository in terms of an URL, and the metadata. In case that the local storage format of the new document differs from the agreed storage format used in the document repositories, the EHR document is transformed prior to its submission. The patient is the only user who is authorized to access the EHR document and who may grant or revoke access to the EHR for other users. This guarantees that patients have full control over their EHRs (as demanded by legislation), and that it is the patients' choice whom they want to share their EHRs with.

Search for documents within the FHIS is performed by (1) passing a query to the document registry, which returns for each matching document the RID together with the URL of the document repository where the document is stored, and (2) retrieving the actual document from the specified document repository. The document registry allows to query for the EHR metadata as well as for the document owner (patient). In order to prevent information leakage, the URL and RID of a matching document are returned by the document registry only to authorized users.

3 Background

Data confidentiality, and thus, privacy can be achieved in different ways. The traditional approach is to explicitly apply access control mechanisms, while disassociation (such as anonymization and pseudonymization) and encryption techniques limit the impact of unauthorized data disclosure. In the following, we shortly discuss these approaches and describe their security implications and limitations for their application in FHIS (cf. Figure 2).

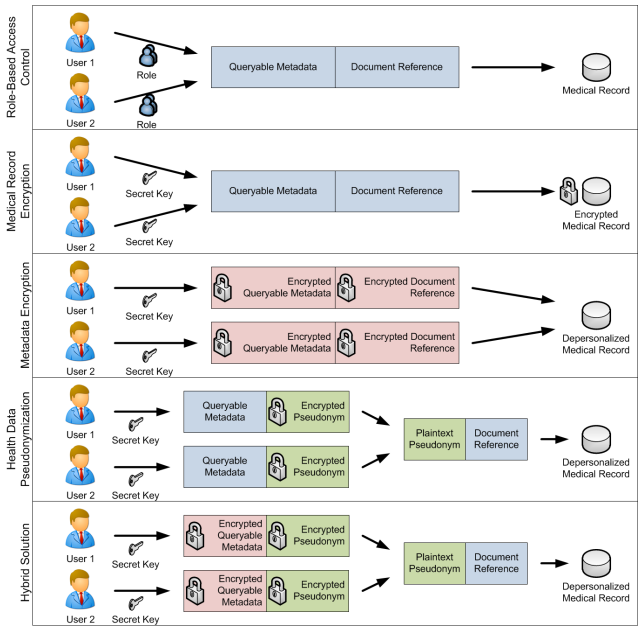

Figure 2. Comparison of Data Protection Strategies

3.1 Traditional Access Control

Current realizations of the FHIS architecture proposed by the IHE, like the national FHIS in Austria called ELGA, rely on role-based access control (RBAC) for data protection where access is granted according to rules expressed in the form of authorization policies (XACML) matching the role the current user embodies. The downside of this model is the central access control module which can be bypassed (cf. [15]) or circumvented by, e.g., administrators with their unrestricted access rights.

3.2 Encryption

The straight forward approach of fully encrypting health records where the patient keeps the decryption key (cf. Figure 2, Medical Record Encryption) is suboptimal due to the potentially large sizes of medical records (especially images) and the limitations for efficient secondary use. Authorizations are also tricky to implement involving either sharing the decryption key (re-encryption required for authorization revocations) or storing a copy of the health record for each user (storage overhead). Therefore, the better alternative to encrypting health documents is the encryption of metadata only, combined with depersonalizing the health records. In this case, the major challenge is how to query within encrypted metadata. Approaches to query within XML data[1] store an XML document as a set of (disjoint) document fragments and use crypto-indexes to facilitate search (cf. [21], [8]). Other approaches store XML documents as single nodes (cf. [9], [18], [7]). In [18], [7], schema information of the metadata XML document is required for query processing where each node is assigned a unique ID according to its path, called path schema ID, and a node instance ID. Queries are expressed in XPath-like expressions and the path looked up in the schema information to determine the corresponding node IDs. In Figure 2 (Metadata Encryption), each authorized user maintains his personal document registry with document metadata and references encrypted with his own secret key. The limitation of this approach is as follows: While data access authorizations can be realized by forwarding the corresponding document metadata and references to the authorized HCPs who incorporate this information in their encrypted metadata storages, the patient no longer has control of deauthorizations, because he lacks direct access to the involved HCPs' secret keys and thus their metadata registries.

3.3 Pseudonymization

Disassociation in e-Health involves the removal of patient-identifying information from the health records. *K-anonymity* refers to releasing data in (equivalence) groups where within each group, the corresponding person's (quasi) identifiers cannot be distinguished from at least k individuals (cf. [16]). This basic concept is extended with approaches such as *l-diversity* [11] and *t-closeness* [10] to further reduce the probability of re-identification.

While anonymity is unreversible and thus its application limited to secondary use (e.g., surveys), a similar technique is pseudonymization but with the difference that it is reversible under specified and controlled circumstances and also keeps data accuracy intact. It is a technique where identifying data is replaced (instead of completely removed) with a specifier (pseudonym) that cannot be associated without knowing a certain secret. Effective pseudonymization requires diligent depersonalization which involves the identification of any patient-identifying data (cf. [5], [22]). In [20], pseudonymization is achieved by first separating the identification data from the anamnesis data which is then stored in a separate database referenced with so called unique *data identification codes* (DIC) as pseudonyms. In [14] and [13], depersonalized health records are assigned so-called *root pseudonyms*, which are only known to the patient, and *shared pseudonyms*, shared between the patient and health professionals as authorization 'tokens'. Knowing the correct pseudonym allows the authorized user to re-link the health record to the corresponding patient.

In Figure 2, pseudonymization is achieved by replacing the document reference with a pseudonym encrypted with both the patient's (user 1) and the HCP's (user 2) keys where the pseudonym is appended to the cleartext document metadata. The document reference is instead assigned to the plaintext pseudonym. The pseudonyms act as document access identifiers where de-authorizations are realized by removing them. The problem with plain pseudonymization is actually the shared cleartext document registry which must not contain any information providing hints that make it possible to re-establish the disconnected patient/document link, especially arbitrary keywords selected by the patient.

4 Hybrid Encryption/Pseudonymization Approach

To overcome the shortcomings of existing approaches, we propose to combine elements of our pseudonymization (cf. [12]) approach and our XML encryption and query scheme (cf. [18], [7]) to a hybrid encryption/pseudonymization (PERiMETER - Pseudonymization and pERsonal METadata EncRyption) approach: As shown in Figure 2 (Hybrid Solution), the depersonalized health records are stored pseudonymized and in cleartext while the searchable metadata is stored encrypted. Metadata encryption ensures that arbitrary keywords that may contain identifying information compromising privacy are protected from unauthorized access. Pseudonymization restores the patient's control of de-authorization by simply deleting the shared pseudonym (mapping), thus severing the logic link between the document reference and the document metadata stored in the de-authorized health care provider's personal document registry.

Figure 3 shows our hybrid concept based on ELGA, the Austrian implementation of a FHIS: The basic infrastructure is composed of a central patient index and a central HCP index, providing demographic and authentication information, including the central (global) patient and HCP identifiers (C-PID and HCP-ID). The health records are managed by distributed and independent XDS affinity domains (ELGA areas) representing a single or a group of HCPs within the same organizational domain. Each of these affinity domains operate their own local patient in-

[1] We concentrate on XML as industry standard for medical metadata (HL7 CDA).

dex with XDS-PIDs (L-PIDs) and a local document registry containing the queryable document metadata including the references to the documents which are stored in the document repositories. Each affinity domain is connected to all other domains and to the central registry via a gateway, communicating with standardized IHE ITI-conforming transactions [1]. The main transactions include the following[2]:

- Central and local patient identifiers are queried for via transactions ITI-45 (PIX Query) and ITI-47 (Patient Demographic Query).
- Document search is realized by ITI-18 (Registry Stored Query) for local search and ITI-38 (Cross Gateway Query) for queries addressed at other affinity domains. Similarly, document retrieval is related to ITI-43 (Retrieve Document Set) and ITI-39 (Cross Gateway Retrieve).
- Providing a new health document involves ITI-41 (Provide and Register Document Set-b) including ITI-42 (Register Document Set-b).

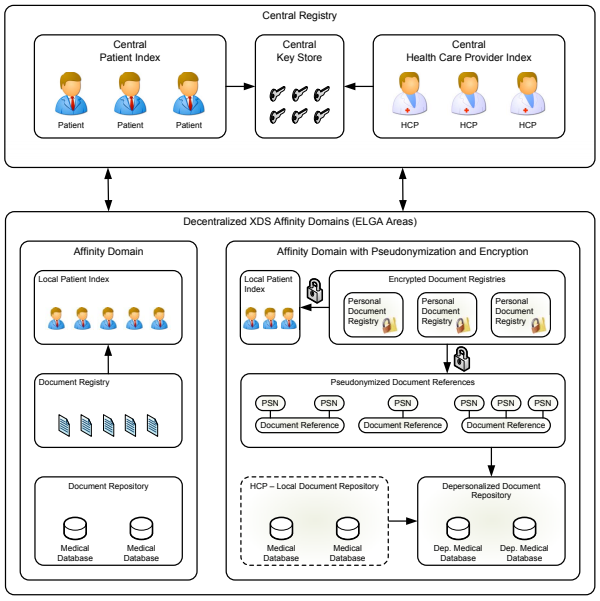

Figure 3. Conceptual Hybrid Architecture

Our hybrid solution may be realized within an affinity domain as depicted in Figure 3: A central key store is introduced keeping the patients' and HCPs' secret keys secured by encryption (cf. Section 5). While the local patient index is left unaltered, the document repository is split such that each FHIS participant (patient whose health documents are stored in the affinity domain, data-providing HCP, and authorized external HCPs that are authorized for data access by the patient) maintains his own personal document repository, encrypted with his own secret key usable only after authentication with a personal security token (i.e., smart card). The document references (in case of ELGA consisting of record identifiers and locations as URLs) originally stored along with the document metadata entries are now replaced with pseudonyms, which are in turn associated with the document references in cleartext in a separate pseudonymized document reference registry. Assuming that the health documents are stored in a standardized format as it is required for interoperability between different independent affinity domains, the documents are depersonalized before they are moved from the HCP's local repository(s) to the depersonalized document repository.

In the following, we describe the data model and element associations of PERiMETER as static view and the main workflows as dynamic view. Thereby, we rely on the notation given in table 1.

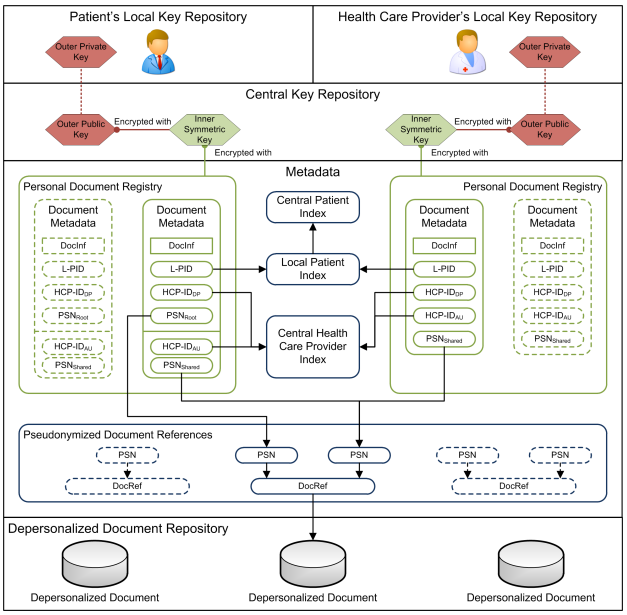

Figure 4. Detailed Static View of the PERiMETER Approach

5 Static View

Figure 4 provides the static view of our hybrid solution including the user-specific cryptographic keys and document/pseudonymization metadata and how these elements are associated to each other.

5.1 Key Repositories

Each user is provided with an outer asymmetric keypair and an inner symmetric key. This set of keys realizes a layered access envelope as follows: the inner symmetric key, which encrypts and decrypts the personal metadata store, is

[2]ITI-8 and ITI-44 transactions as shown in Figure 1 are not considered as they do not play a direct role in our approach.

$C\text{-}PID, L\text{-}PID$	central and local (domain-specific) patient identifier
$HCP\text{-}ID$	central HCP identifier
$OPK/OPuK$	outer private and public key, private key persisted on token
ISK	inner symmetric key
$\{item\}key$	item encrypted with key
$[item]^+$	one or more items in a set
$[item]^*$	none, one, or more items in a set
U, P, DP, AU	user (patient or HCP), patient, document-providing and authorized HCP
PSN_{Root}	root pseudonym only known to patient
PSN_{Shared}	shared pseudonym known to both patient and HCP representing an access authorization
Doc	actual health document
$DocInf$	document information stored in personal document registry including document type, format, date and time, ...
$DocRef$	document reference with record identifier (RID) and location (URL)
$query$	query elements

Table 1. Notation

encrypted with the outer public key. The central key repository contains the following elements:

$$[C\text{-}PID\ or\ HCP\text{-}ID, \{ISK\}\, OPuK, OPuK] \tag{1}$$

The outer private key is stored within the secured confinement of the user's smart card only (local key repository) protected by a PIN:

$$[C\text{-}PID\ or\ HCP\text{-}ID, OPK] \tag{2}$$

Thus, both the smart card as well as the PIN are required to decrypt the inner symmetric key to gain access to the personal document registry. In case of a lost or damaged smart card, the inner symmetric key would be lost for good; therefore a backup mechanism is required. For increased protection against misuse, we propose to use threshold-based secret sharing schemes (e.g., [19]) to distribute shares of the inner symmetric key to different share holders (e.g., relatives, health care providers, administrators).

5.2 Personal Document Registry

Each patient's personal document registry entry includes the document information (document type, formatting information, date and time, etc.), as well as the local patient identifier and the document provider's identifier, all encrypted with the patient's inner symmetric key. Root pseudonyms act as primary document access identifiers for patients, while shared pseudonyms and authorized HCPs' identifiers represent individual document access authorizations.

$$[\{DocInf, L\text{-}PID, HCP\text{-}ID_{DP}, PSN_{Root}, [HCP\text{-}ID_{AU}, PSN_{Shared}]^*\}\, ISK_P] \tag{3}$$

The authorized HCP's document registry stores practically the same elements as the patient's, but with the exception of the root pseudonym (only known to the patient) and of course the authorizations for other HCPs.

$$[\{DocInf, L\text{-}PID, HCP\text{-}ID_{DP}, HCP\text{-}ID_{AU}, PSN_{Shared}\}\, ISK_{AU}] \tag{4}$$

5.3 Pseudonymized Document References

The remaining elements are the pseudonyms (exactly one root pseudonym and a shared pseudonym for each individual authorization for this health document) mapped to the document references (unique identifier and URL) stored in cleartext.

$$\left[DocRef, [PSN]^+\right] \tag{5}$$

6 Dynamic View

As a common precondition to the following workflows, the user's inner symmetric key has to be available. During authentication at an identity provider involving the user's outer keypair (e.g., nonce-based challenge/response), the central key repository is accessed to retrieve the encrypted inner symmetric key which is transferred to the user's smart card. With the outer private key located at the smart card, the inner symmetric key is decrypted and remains at the card as long as it is needed, i.e., as long as the current session is active, and is automatically erased when the smart card is removed from the card reader. The inner symmetric key never leaves the card in an unencrypted state. The smart card therefore acts as a temporary secure keystore for the inner symmetric key, as well as a hardware-based cryptographic device.

The workflow steps are modeled in UML sequence diagrams, emphasizing the messages exchanged between the entities described in the previous section.

6.1 Data Retrieval

Data retrieval involves querying the personal document registry, retrieving and selecting the desired pseudonym, and forwarding the pseudonym to the pseudonymized document reference storage to finally acquire the actual health record via the document reference. The workflow is basically the same for both patient and HCP. The only difference is that the patient queries for his root pseudonyms, while the authorized HCP relies on the shared pseudonyms.

1. The user (patient or authorized HCP) formulates the query and sends it to the personal document registry. The query (as XPath-like expression) is processed, and the registry returns any matching document metadata including the pseudonym in encrypted form.

2. The user selects the desired pseudonym(s) inspecting the document metadata information and sends the pseudonym(s) to the document reference storage

which forwards the associated document reference(s) to the document repository to return the corresponding health record(s).

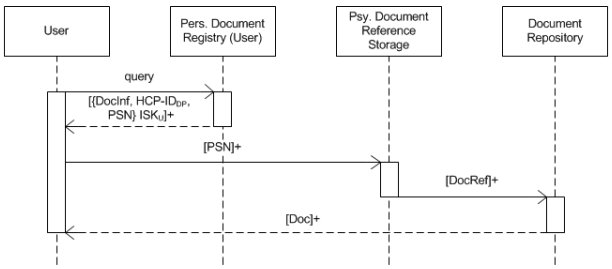

Figure 5. Data Retrieval

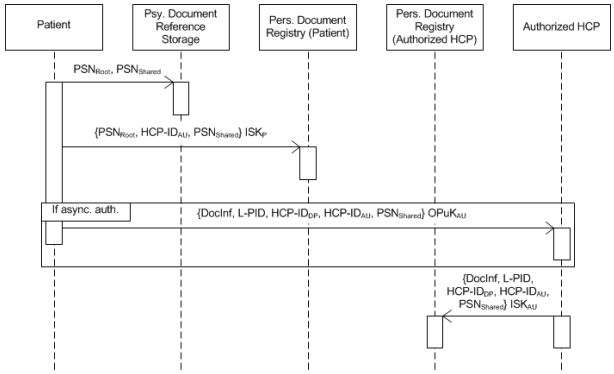

Figure 6. Authorization

6.2 Authorization

A data access authorization is created for an individual health record and is realized by a randomly selected new shared pseudonym. The pseudonym is shared between the patient and the HCP to be authorized and referenced with the particular document. In addition, the corresponding document metadata is retrieved from the patient's personal document registry and copied to the HCP's personal registry.

1. First the patient executes the steps described in the previous section to retrieve the health document's root pseudonym and metadata. Then the patient randomly selects a new shared pseudonym and transfers it to the document reference storage, where it is associated with the same document reference as the root pseudonym. Furthermore, the patient updates[3] his personal document registry with the authorization information using the root pseudonym as metadata 'identifier'.
   (a) When both patient and HCP are present at the same machine, the document info, shared pseudonym, and patient's and HCPs' identifiers are simply re-encrypted and stored in the (authorized) HCP's personal document registry. This can be referred to as synchronous authorization.
   (b) If not, the patient retrieves the HCP's outer public key from the key repository via HCP-ID and sends the metadata elements as notification to the HCP (e.g., via a centrally accessible notification storage area). If the HCP then logs into the FHIS, he retrieves this notification, re-encrypts the elements and appends them as new document metadata entry in his personal document registry. This is referred to as asynchronous authorization.

[3] Similar to a query, the update or insertion process of elements into the encrypted document registries involves multiple individual transactions.

6.3 Document Storage

Adding a new health record requires multiple individual steps: Assuming that the document is created and stored in the HCP's local document repository, it needs to be depersonalized before being stored in the FHIS document repository. Then the patient has to be informed of the new record to be available, e.g., in the form of a notification including the document metadata extracted from the health record and pseudonymization metadata. When the patient logs in, the system updates the patient's personal document registry with the new entry.

1. First, the document provider copies the new document from his local HCP repository to the depersonalized FHIS document repository after removing any patient-identifying details. In addition, a randomly selected shared pseudonym is sent along to be associated with the document reference generated by the repository and forwarded to the document reference storage.

2. The document provider retrieves the patient's outer public key to forward the document information (probably automatically extracted from the health document and extended with arbitrary keywords), his HCP and patient identifiers, and the pseudonym in the form of a notification to the patient. The provider also stores the document metadata in his personal document registry and is automatically authorized for data access.

3. Upon logging in, the patient retrieves the notification and decrypts the elements with his outer private key. In addition, the patient also creates a new root pseudonym to be appended to the document reference (via shared pseudonym). Then the patient registers the document metadata along with the authorization information concerning the document provider in his personal document registry, concluding the document storage procedure.

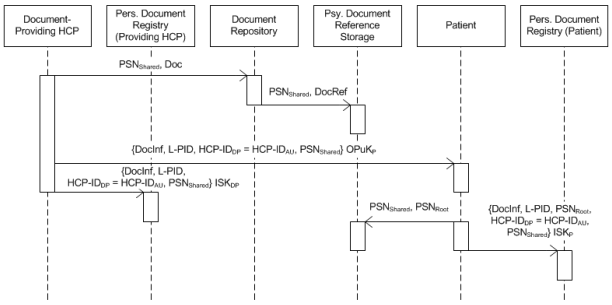

Figure 7. Document Storage

7 ITI Transaction Modifications

The integration of our hybrid approach requires certain modifications of the ITI transactions mentioned earlier in this article. Generally speaking, queries need to be reformulated and require multiple lookups, while result sets contain pseudonyms instead of explicit document references; document storage requires changes to account for the pseudonyms. In particular, the following modifications are needed:

- ITI-18: Reformulation of queries to suit the XPath format as described in [7].
- ITI-43: Modification to accept one or more pseudonyms instead of document references which are not transferred to the document consumer.
- ITI-39: Similar to ITI-43, modification to accept one or more pseudonyms instead of document references.
- ITI-41: Document storage in the document repository only, i.e., without automatic registration/metadata storage in the document registry (ITI-42).
- ITI-42: Modification of document registration to include pseudonym mapping storage as well as updating the personal document registries

8 Benefits and Limitations

Summarized, PERiMETER has the following benefits:

- Encryption of document metadata effectively hides the relationship between health documents and patients, unless authorized, thus rendering database theft (dump) useless.
- The personal registries provide different 'views' on the entirety of the health records at a need-to-know basis.
- Authorizations are defined on a fine-grained level (discretionary access control).
- Pseudonymization allows deauthorizations without the authorized HCP's consent, giving the patient full control.
- Encryption and decryption operations are executed at the smart card only, minimizing the risk of keys being compromised.
- The encrypted metadata and pseudonymized health records protect the patients' privacy even against internal attackers (e.g., administrators) who lack access to the inner symmetric keys required for decryption and de-pseudonymization.

Still, we identified certain limitations and challenges that need to be solved:

- Integration with the IHE standard remains a major challenge, mainly because of the different authorization, access control, and logging strategies (ATNA, BPPC, etc.), partly contradicting with the ideas of security by encryption and pseudonymization. Furthermore, the modifications of IHE-conforming transactions as required for the metadata query process affect interoperability.
- Redundant metadata storage for each user, unless the encrypted document metadata is shared with each authorized party. This alternative requires that each de-authorization includes re-encryption and redistribution of the new crypto key.
- Communication overhead of the query process, probably involving multiple transactions and crypto operations. Considering the limited number of different queries (stored query), this issue can be considerably reduced by creating customized secondary index structures and suitable data fragmentation (cf. [6]), thereby reducing the amount of necessary individual transactions.

9 Conclusion

Traditional data protection techniques like role-based access control are secure as long as they are not circumvented by internal attackers. In FHIS, a potential internal attacker may get access to sensitive health data. Therefore, this paper proposed a hybrid data protection approach integrating pseudonymization with metadata encryption which allows health documents to be stored unencrypted and thus be available for secondary use, while the metadata is encrypted but still searchable. This solution also permits fine-grained and patient-controlled authorization for individual documents.

Acknowledgments

This work was supported by grants of the Austrian Government's BRIDGE Research Initiative (contract 824884),

the FIT-IT Research Initiative (contract 816158) and was performed at the research center Secure Business Austria funded by the Federal Ministry of Economy, Family and Youth of the Republic of Austria and by the City of Vienna.

References

[1] IHE IT Infrastructure (ITI) Technical Framework 7.0. Technical report, Integrating the Healthcare Enterprise (IHE), August 2010.

[2] R. C. Barrows and P. D. Clayton. Privacy, confidentiality, and electronic medical records. *Journal of the American Medical Informatics Association*, 13:139–148, 1996.

[3] F. R. Ernst and A. J. Grizzle. Drug-related morbidity and mortality: Updating the cost-of-illness model. *Journal of the American Pharmacists Association*, 41(2):192–199, 2001.

[4] S. Fischer-Hübner. *IT-Security and Privacy: Design and Use of Privacy-Enhancing Security Mechanisms*. Springer, Berlin, 2001.

[5] F. Friedlin and C. Mcdonald. A software tool for removing patient identifying information from clinical documents. *Journal of the American Medical Informatics Association*, 15:601–610, 2008.

[6] K. Grün. A generic framework for querying and updating secondary XML index structures. In *Proceedings of SIGMOD2007 Ph.D. Workshop on Innovative Database Research 2007 (IDAR2007)*, 2007.

[7] K. Grün, M. Karlinger, and M. Schrefl. Schema-aware labelling of XML documents for efficient query and update processing in SemCrypt. *Comput. Syst. Sci. Eng.*, 21(1), 2006.

[8] R. C. Jammalamadaka and S. Mehrotra. Querying Encrypted XML Documents. In *Proceedings of the 10th International Database Engineering and Applications Symposium*, pages 129–136. IEEE Computer Society, 2006.

[9] J.-G. Lee and K.-Y. Whang. Secure query processing against encrypted XML data using Query-Aware Decryption. *Inf. Sci.*, 176(13):1928–1947, 2006.

[10] N. Li, T. Li, and S. Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *IEEE 23rd International Conference on Data Engineering (ICDE2007)*, pages 106–115, 2007.

[11] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. L-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data*, 1(1):3, 2007.

[12] T. Neubauer and J. Heurix. A methodology for the pseudonymization of medical data. *International Journal of Medical Informatics*, 2010.

[13] B. Riedl, V. Grascher, S. Fenz, and T. Neubauer. Pseudonymization for improving the privacy in e-health applications. In *Proceedings of the Forty-First Hawai'i International Conference on System Sciences*, page 255, 2008.

[14] B. Riedl, T. Neubauer, and O. Boehm. Patent: Datenverarbeitungssystem zur Verarbeitung von Objektdaten. *Austrian Patent, Nr. 503291, September*, 2007.

[15] R. Russell, D. Kaminsky, R. F. Puppy, J. Grand, D. Ahmad, H. Flynn, I. Dubrawsky, S. W. Manzuik, and R. Permeh. *Hack Proofing Your Network (Second Edition)*. Syngress Publishing, 2002.

[16] P. Samarati and L. Sweeney. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical report, SRI Int'l, 1998.

[17] T. Schabetsberger, E. Ammenwerth, G. Göbel, G. Lechleitner, R. Penz, R. Vogl, and F. Wozak. What are functional requirements of future shared electronic health records? In R. Engelbrecht, A. Geissbuhler, C. Lovis, and G. Mihalas, editors, *European Notes in Medical Informatics: Connecting Medical Informatics and Bio-Informatics; MIE2005*, pages 1070–1075, 2005.

[18] M. Schrefl, K. Grün, and J. Dorn. SemCrypt - Ensuring Privacy of Electronic Documents Through Semantic-Based Encrypted Query Processing. In *21st International Conference on Data Engineering Workshops (ICDEW'05)*, page 1191. IEEE Computer Society, 2005.

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

[20] C. Thielscher, M. Gottfried, S. Umbreit, F. Boegner, J. Haack, and N. Schroeders. Patent: Data processing system for patient data. *Int. Patent, WO 03/034294 A2*, 2005.

[21] Y. Yang, W. Ng, H. L. Lau, and J. Cheng. An Efficient Approach to Support Querying Secure Outsourced XML Information. In *CAiSE*, Lecture Notes in Computer Science, pages 157–171. Springer, 2006.

[22] R. Yeniterzi, J. Aberdeen, S. Bayer, B. Wellner, L. Hirschman, and B. Malin. Effects of personal identifier resynthesis on clinical text de-identification. *Journal of the American Medical Informatics Association*, 17:159–168, 2010.