Economic and Security Aspects of the Appliance of a Threshold Scheme in e-Health

Bernhard Riedl*, Veronika Grascher*, Mathias Kolb*, Thomas Neubauer*
*Secure Business Austria, Vienna
Email: {riedl, grascher, kolb, neubauer}@securityresearch.ac.at

***Abstract*—Today, the healthcare sector is driven by the need to reduce costs while simultaneously increasing the service quality for patients. This goal can be reached by the implementation of an EHR (Electronic Health Record) system. Several architectures have been proposed, but lack appropriate security mechanisms to protect the patients' privacy. In this publication we outline our approach PIPE (Pseudonymization of Information for Privacy in e-Health), which is applicable for the primary and secondary usage of health data and give insights on the security of our technique. Further we state the economic constraints, by proposing a threshold scheme to secure the tokens needed for accessing the system.**

I. Introduction

Nowadays, it is a demanding goal for operators of health care systems to handle the increasing number of patients. The electronic health record (EHR) is one possibility to get the resulting costs under control [1]–[3]. Moreover, medical suppliers offer an abundant supply of medication, which can hardly be handled by health care providers (HCP). Therefore, decision support systems on drug interaction are needed [4]–[7] to assure a better quality of patients' treatment. Further, the implementation of standard processes within the EHR and accordant workflows, may also help to increase the medical service quality [8], [9].

Besides the fact that it is still a challenge for nation-wide health care systems to harmonize their processes and exchange data in standardized formats (for example HL7 [10]), the security of proposed EHR architectures [11]–[17] have to be observed as a matter of fact [18], [19]. The EHR may offer better security as the traditional paper-based records system [20], if the privacy of the individual is protected accurately. Firstly, the patient should be the person with security clearance [21]. In other words, the patient should be in full control of her data, which means that she decides about access and authorization to her data. Secondly, as life-long sensitive medical information is stored within an EHR system, this information needs to be protected. For example a history about substance abuse or HIV-infection could lead to denying medical insurance coverage [20].

One possibility to protect the patients' medical records is to encrypt the information. Unfortunately, as for example radiology images tend to be very large [22], [23] and encryption is a time-consuming operation, such approaches are most of the time not applicable. Pseudonymization is a technique to conceal the relation between a patient and her medical data without the need for encryption [18], [19], [24]–[27], which is typically only used for the secondary usage of anamnesis data (cf. [16], [17], [28]).

To circumvent security flaws against an EHR system we developed PIPE (Pseudonymization of Information for Privacy in e-Health) [18], [19], [25], [26], which is based on a novel pseudonymization approach. It can be used as stand-alone or add-on to existing architectures. As we base our security model on keys held on security tokens (in our prototype smart cards), there is the need to assure the availability of these keys, even if a patient looses access to her security token. In this publications we outline the security model of our approach and give economic insights on the applied backup-mechanism for lost or destroyed smart cards.

II. Related Work

As in a typical relational database, in a medical database, a relation between the identification data is separated from the anamensis data by mapping the primary key of the identification table to the foreign key of the anamnesis table. If we would not store the foreign key in the anamnesis table and someone does not know this relation, it would not possible for her to associate a certain anamnesis with a specific individual. Hence, the goal of a pseudonymized medical storage is to conceal this relation [13], [16]–[19], [25], [26] and thus to protect the patients' privacy.

To hide the association between the identification data and the medical data, the foreign keys of the anamnesis table are not deleted, but substituted with pseudonyms [18], [19], [24]–[27]. The calculation of these pseudonyms is based on a secret-key algorithm. Therefore, only the instances who know the appropriate pseudonymization key are able to re-establish the relation between a patient and her anamnesis.

Pommerening proposed a system for the secondary usage of medical data [16], [17] where a pseudonym is formed by a combination of a hashing and an encryption technique with the key kept centralized inside the system. This method poses a major security threat for the patients' medical data because persons working for such a system might be attacked for example with social engineering [20], [29], [30]. Moreover, the dependency of this approach on a (study-)life-long pseudonym for every patient, which was recently named a person's identifier by the Article 29 Working Party of the European Union[1] [31], could lead to data mining [32] or

[1]http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2007/wp136_en.pdf

profiling attacks. The only reliable approach is that for every combination of patient, health-care-provider and anamnesis, a single pseudonym exists, which cannot be related with any participating actor.

From a security point of view, the system of Thielscher et al. [13] is comparable to Pommerening's approach. Indeed, Thielscher uses smart cards for the patients as security tokens, which produce the necessary pseudonyms for data access. In case a smart card is lost and the key on this security token needs to be restored, they equip their system with a pseudonymization computer, on which backups are held. To decrease the risks of hacking attacks from outside, they operate this computer offline. Nevertheless, this technique still makes insider attacks possible, for example by bribing persons inheriting administrative roles. Thus, there is the need to avoid unauthorized access to these backup-pseudonyms, respectively to these backup-pseudonymization keys. A threshold scheme for secret sharing [33] may be applied to divide the backup keys between more administrators.

Besides the secure creation of pseudonyms, with the constraint of decentralized keys on security tokens such as smart cards and avoidance of life-long pseudonyms, a secure authorization technique needs to be established for an EHR. One possibility is to share the patient's 'root-key', but in that case revoking of authorizations would hardly be possible because this key and accordantly all pseudonyms have to be changed to deny access to a revoked instance. Thus, other keys have to be created in order to establish a reliable authorization. Schmidt [34] describes a system, which is mainly based on a public key infrastructure (PKI) in combination with encryption. This approach provides every participant with different keys, which establish access to a subset of the fully or partially encrypted data. As aforementioned, encryption is a time-consuming operation and medical data tend to be very large, for example an x-ray consumes 6 MB, a mammogram 24 MB [22] or a computer tomography scan counts up to hundreds of MB [23]. Thus, it is not possible to encrypt for example radiology images.

In the next section we introduce our approach PIPE, which is solely based on pseudonymization instead of encryption. The access to the anamnesis for primary and secondary usage is handled through a security hull structure, which allows granting authorizations by sharing encrypted secrets. Compared to the encryption of the medical data itself, the encryption of secrets like keys or relations is fast enough to meet the needs of a health care team. Moreover we provide a secure fall-back mechanism if a system participant looses access to her security token. We give economic and statistic security insights on our applied technique.

III. System Overview

The goal of our architecture [18], [19], [25], [26] is to gain the optimal trade-off between security, usability and performance.

In table 1 we provide an overview of the keys and abbreviations used in our system. Note, that all private keys (where K stands for key) are identified as K^{-1} (e.g., the patient's inner private key will be named $\widehat{K}_A^{-1}$).

As sketched in figure 1, PIPE is based on a security hull-architecture [18], [19], [25], [26]. In the most outer layer — the user permissions layer — every user U 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 — the 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 ψ_{i_0}. All other pseudonyms ψ_{i_j} are disjunct for any patient, health-care-provider and anamnesis combination. If, for example, two health care providers have been authorized for a specific anamnesis φ_i, three pseudonyms (ψ_{i_0}, ψ_{i_1} and ψ_{i_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 $\overline{K}_A$ or any authorized health care provider's C inner symmetric key $\overline{K}_C$. To get access to these particular keys, the authorized users' inner private key has to be used. We provide a formal depiction of a patient's encrypted anamnesis storage in equation 1 and similar for a health care provider in equation 2.

$$\left\{\left\{\left\{\{\psi_{i_0} \mapsto \varphi_i\}_{\overline{K}_A}\right\}_{\widehat{K}_A}\right\}_{K_A}\right\} \tag{1}$$

$$\left\{\left\{\left\{\{\psi_{i_j} \mapsto \varphi_i\}_{\overline{K}_C}\right\}_{\widehat{K}_C}\right\}_{K_C}\right\} \tag{2}$$

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 $\widehat{K}_A^{-1}$. Firstly, a relative B may encrypt the patient's inner private key with her inner public key $\widehat{K}_B$. Thus, she is also able to decrypt the patient's inner symmetric key until the patient changes it. We show these encryption envelopes in equation 3.

$$\left\{\left\{\left\{\left\{\left\{\{\psi_{i_0} \mapsto \varphi_i\}_{\overline{K}_A}\right\}_{\widehat{K}_A}\right\}_{\overline{K}_B}\right\}_{\widehat{K}_B}\right\}_{K_B}\right\} \tag{3}$$

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 [33] between a set of operators $\mathcal{O}$, who are randomly assigned to hold a part of the users' secrets. Due to financial reasons, we propose a combination of human operators $\mathcal{H}$ and machine operators $\mathcal{M}$, for example Hardware

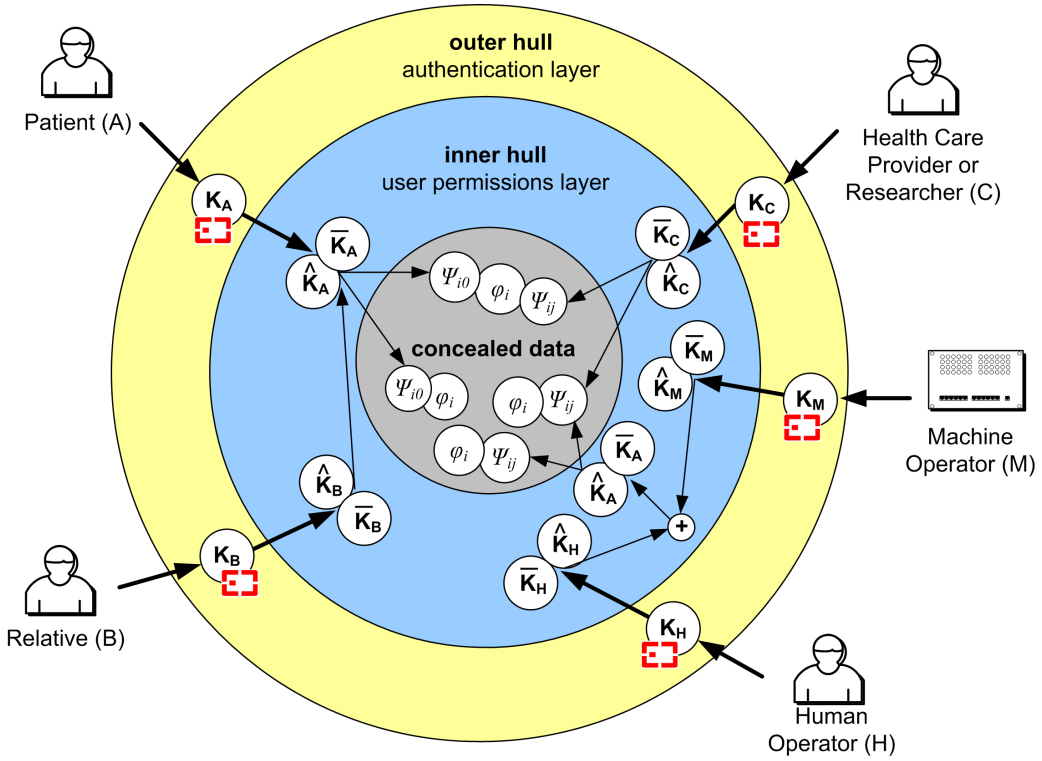

Fig. 1. Layered model representing the authorization mechanism [18], [19], [25], [26]

	Patient	*Relative*	*HCP*	*Operator*	*Human O*	*Machine O*	*Logic*
abbreviation	A	B	C	O	H	M	L
unique identifier	A_{id}	B_{id}	C_{id}	O_{id}	H_{id}	M_{id}	
(outer public key, private key)	(K_A, K_A^{-1})	(K_B, K_B^{-1})	(K_C, K_C^{-1})	(K_O, K_O^{-1})	(K_H, K_H^{-1})	(K_M, K_M^{-1})	
(inner public key, private key)	$(\widehat{K}_A, \widehat{K}_A^{-1})$	$(\widehat{K}_B, \widehat{K}_B^{-1})$	$(\widehat{K}_C, \widehat{K}_C^{-1})$	$(\widehat{K}_O, \widehat{K}_O^{-1})$	$(\widehat{K}_H, \widehat{K}_H^{-1})$	$(\widehat{K}_M, \widehat{K}_M^{-1})$	
inner symmetric key	$\overline{K}_A$	$\overline{K}_B$	$\overline{K}_C$	$\overline{K}_O$	$\overline{K}_H$	$\overline{K}_M$	K_L
key share				$\sigma_\iota(K)$	$\sigma_{\mathcal{H}_\iota}(K)$	$\sigma_{\mathcal{M}_\iota}(K)$	
medical data / anamnesis	φ_i						
pseudonym	ψ_{i_j}						

TABLE I
DEFINITION OF SYSTEM ATTRIBUTES

Security Modules (HSMs) [35]. For a HSM, we recommend FIPS Level 3, which has physical security mechanism to prevent the access to critical security parameters [36].

In the following sections, we provide insights on our backup keystore.

IV. Establishing a Secure Backup Keystore

Since users need a fall-back mechanism in case they have lost their smart cards, operators $\mathcal{O}$ hold the users' inner private keys on behalf of the patients. To avoid misuse of these rights, we applied Shamir's threshold scheme [33] to split the secret into several parts [18], [19], [25], [26]. Following Shamir, two parameters can be defined for sharing a secret, (i) the number of issued 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 that is needed to re-establish a shared secret, the higher the level of security, assuming the operators are randomly assigned, each holding one share of a certain secret.

Certainly, 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 $\mathcal{H} \subset \mathcal{O}$ and machine operators $\mathcal{M} \subset \mathcal{O}$. In this publication we state the condition that both human and machine operators are required for recovering a secret, which is guaranteed by applying a two-folded variant of Shamir's secret sharing scheme. In figure 2, we show how a patient's inner private key $\widehat{K}_A^{-1}$ is divided amongst the human $\mathcal{H}$ and machine operators $\mathcal{M}$. Firstly, we divide the patient's inner private key into two parts, $\sigma_{\mathcal{H}}(\widehat{K}_A^{-1})$ for the human operators and $\sigma_{\mathcal{M}}(\widehat{K}_A^{-1})$ for the machine operators. Afterwards, the threshold scheme is again applied to subdivide $\sigma_{\mathcal{H}}(\widehat{K}_A^{-1})$ into the number of n_H assigned human operators. Analogous $\sigma_{\mathcal{M}}(\widehat{K}_A^{-1})$ is distributed amongst n_M assigned

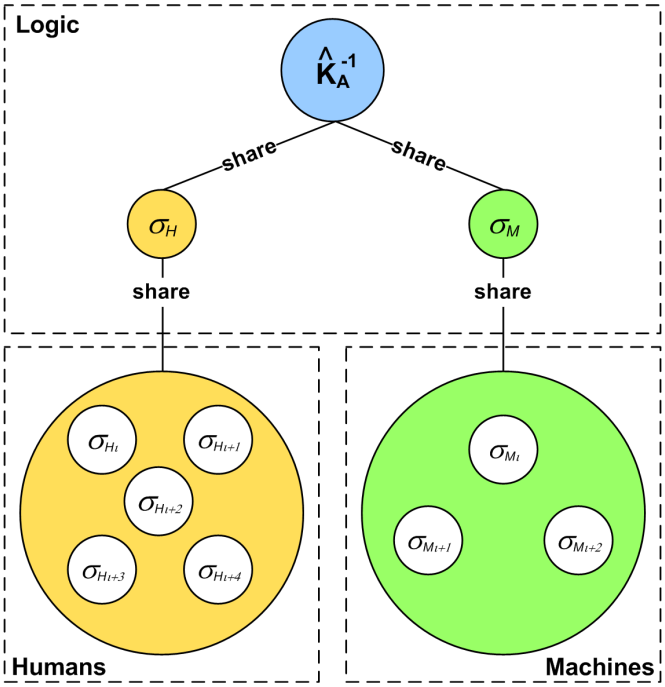

Fig. 2. Two-folded variant of Shamir's threshold scheme

machine operators. Finally, PIPE encrypts the shares σ_{H_ι} and σ_{M_ι} with a system's key and sends them to the particular assigned operators. Using encryption helps to conceal the relation between an operator and a patient. If an operator has been successfully bribed, the attacker only gains access to her part of a user's shared secret, but not to the related parts of the other participants. The delta between the number of necessary operators k and assigned operators n serves as major availability constraint in our system, because a human operator may be ill or a machine operator could have a malfunction.

The following example provides an overview of the costs, which we split in initial costs $C_{initial}$ (cf. Equation 4) and current costs $C_{current}$ (cf. Equation 9). The latter is based on equation 6, which is used for calculating the estimated key recovering requests per anno and the equations 7 and 8 for the numbers of human as well as machine operators. Moreover, we state the time to set-up the system, in other words the initial time $t_{initial}$ (cf. Equation 5). We define the necessary parameters for our obligations in table II.

$$C_{initial} := U * \left(\frac{n_H * C_H}{r_H^i} + \frac{n_M * C_M}{r_M^i} \right) \tag{4}$$

$$t_{initial} := \begin{cases} \frac{n_H * U}{r_H^i} \\ \frac{n_M * U}{r_M^i} \end{cases} \tag{5}$$

$$r := \frac{U}{sc} + U * \frac{p}{100} \tag{6}$$

$$|\mathcal{H}| := \begin{cases} \frac{r * k_H}{r_H^c} & \text{if } \frac{r * k_H}{r_H^c} \geq n_H \\ n_H & \text{else} \end{cases} \tag{7}$$

abbr.	description
k_H	necessary human operators
k_M	necessary machine operators
n_H	assigned human operators
n_M	assigned machine operators
sc	smart card lifetime
p	percentage of lost smart cards per year
r	estimated requests per year
r_H^i	manageable requests of one human operator at set-up
r_M^i	manageable requests of one machine operator at set-up
r_H^c	manageable requests of one human operator per year
r_M^c	manageable requests of one machine operator per year
C_H	costs of one human operator including overhead
C_M	accumulated prime and maintenance costs of one machine operator per year

TABLE II
DEFINITION OF ATTRIBUTES USED IN CALCULATIONS

$$|\mathcal{M}| := \begin{cases} \frac{r * k_M}{r_M^c} & \text{if } \frac{r * k_M}{r_M^c} \geq n_M \\ n_M & \text{else} \end{cases} \tag{8}$$

$$C_{current} := |\mathcal{H}| * C_H + |\mathcal{M}| * C_M \tag{9}$$

We assume an EHR system for 50 Mio. users, which would depict the population of England [37]. A typical human operator with adequate education and experience would earn about Euro 36.000 [38] per year. The overhead costs, which will occur for example by working place expenses or equipment, result in 40 percent surplus. Hence, the total costs for one human operator C_H would be approximately Euro 50.000 per year. An average human works 200 days a year [39], which results in about 1.600 working hours on full-time employment. A human needs about 30 seconds to control a case of a lost smart card. Thus, she is able to contribute 192.000 requests r_H^c a year to recover inner private keys $\widehat{K}_A^{-1}$. Note that this number does not include the identification task and only refers to one of the necessary human operators $\mathcal{H}^k$.

As we already mentioned, it is possible to add machine operators M to the system to decrease the operational costs. The total costs C_M for these machines can be split into prime costs divided by lifetime and maintenance costs per year. We assume that the prime costs, the implementation costs and the running costs are Euro 10.000 each for an appropriate HSM with a lifetime of 10 years[2,3]. This results in Euro 3.000 for a machine operator per year including all overhead costs. In our case a HSM is able to handle 360.000 operations per hour, which means that it could conduct 3.150.446.400 requests r_M^c a year in best case — calculated with an uptime of 99.9 percent. We assume that this number is equivalent at set-up time. During the initiation of the system, all assigned human operators have to encrypt the key shares $\sigma_{H_\iota}(\widehat{K}_A^{-1})$ and the machine operators the key shares $\sigma_{M_\iota}(\widehat{K}_A^{-1})$ of every

[2]nCipher assumes a lifetime of 14 years and initial costs of Euro 12.700 for a nShield PCI 500, this device is able to handle 500 requests per second. Online: http://www.ncipher.com

[3]Utimaco assumes a lifetime of 10+ years and initial costs of USD 9.600 for a Safeguard SecurityServer S10, this device is able to handle 100 request per seconds Online: http://www.utimaco.us

participating user. Opposite to that thesis, human operators are able to conduct more requests at set-up time compared to the yearly manageable requests because the encryption of the secret shares can be done in bulk. Therefore, these requests are only limited by the smart cards' runtime, which is not more than 1 second per operation. This leads to a total number of 5.760.000 manageable requests r_H^i on set-up for human operators. Regarding the smart card constraints, the typical time to live is 5 years and we assume that the loss rate counts up to approximately 7 percent per year.

Shamir stated, that a minimum of $n = 2k - 1$ users, in our case operators, is required to re-calculate a certain secret, which makes a "very robust key management scheme" [33]. A system with 5 assigned/3 necessary human and 3 assigned/2 necessary machine operators as well as the constraints defined above can be handled by 211 human operators and 3 machine operators. Nevertheless, we use 5 machine operators, which will only result in additional costs of Euro 6.000, because with 2 assigned and 3 necessary machine operators an attacker would know that every M has to hold a share of a certain user. Thus, in our example with $|\mathcal{H}| = 211$ and $|\mathcal{M}| = 5$, the initial costs are Euro 2.170.282 which is equal to Euro 0,043 per smart card. The current costs are Euro 10.565.000 per year or Euro 0,783 per worn-out, destroyed, stolen or lost smart card.

The initial set-up for our system comprises a two-folded process and lasts the time which is necessary to divide and distribute the users' key shares amongst the assigned human and machine operators. Hence, 250.000.000 shares for the human operators and 150.000.000 machine operator requests have to be handled by the total number of human and machine operators, if the shares will be randomly distributed. As already mentioned, the amount of manageable shares diverges between the current and inital manageable operations. Latter are used to allocate the occuring requests to the necessary time. Note, that in case the processes of the human and machine operators are started and run concurrently, the maximum of both time parameters results in the total initial time of $\approx 41,14$ working days in our example.

In the next section we outline security aspects of our example.

A. Security Investigations on the Backup Keystore

Following Shamir [33], it is not possible to compute the user's inner private key by combining $k-1$ shares. We defined the number of bribed human operators with b_H and the number of bribed machine operators with b_M. If an attacker is able to bribe $b_H \geq k_H$ or $b_M \geq k_H$ operators, she may succeed in unveiling one of the subsecrets $\sigma_{\mathcal{H}}$ or $\sigma_{\mathcal{M}}$. Equation 10 states the probability of guessing at least the set of necessary human or machine operators for a specific user under the condition, that the operators do not know for whom they are holding shares.

$$P(k \leq X \leq n) = \sum_{\iota=k}^{n} \frac{\binom{n}{\iota}\binom{|\mathcal{O}|-n}{b-\iota}}{\binom{|\mathcal{O}|}{b}} \tag{10}$$

As we implemented the threshold scheme as two-folded process, an attacker is not able to reconstruct the user's inner private key until both subsecrets are successfully computed. Bribing $b_H \geq k_H$ human operators does not influence the probability of recovering the machine operators' subsecret $\sigma_{\mathcal{H}}$, too. Therefore, we are able to define these two events as statistically independent. Following the multiplication rule for independent events, the probability for their intersection, which means combining all necessary shares of a specific secret, is equivalent to the product of the single probabilites.

bribed	$P(\sigma_{\mathcal{H}})$	$P(\sigma_{\mathcal{M}})$	$P(\sigma_{\mathcal{H}} \cap \sigma_{\mathcal{M}})$
$b_H = 3, b_M = 2$	$<$ 0,00001	0,3	$<$ 0,00001
$b_H = 3, b_M = 3$	$<$ 0,00001	0,7	$<$ 0,00001
$b_H = 4, b_M = 2$	0,00003	0,3	$<$ 0,00001
$b_H = 4, b_M = 3$	0,00003	0,7	0,00002
$b_H = 5, b_M = 2$	0,00006	0,3	0,00002
$b_H = 5, b_M = 3$	0,00006	0,7	0,00004
$b_H = 10, b_M = 2$	0,00074	0,3	0,00022
$b_H = 10, b_M = 3$	0,00074	0,7	0,00052
$b_H = 20, b_M = 2$	0,00651	0,3	0,00195
$b_H = 20, b_M = 3$	0,00651	0,7	0,00456
$b_H = 30, b_M = 2$	0,02144	0,3	0,00643
$b_H = 30, b_M = 3$	0,02144	0,7	0,01501

TABLE III
DIFFERENT COMBINATIONS OF BRIBED H AND M OPERATORS

Table III provides a security overview of different combinations of bribed human and machine operators with the parameters of the example from the previous section, k_H=3, n_H=5 for $|\mathcal{H}| = 211$ human operators and k_M=2, n_M=3 for $|\mathcal{M}| = 5$ machine operators.

The first column shows the single probability of recovering the subsecret σ_H. This data can also be interpretated as security investigations on a single-folded process without the application of machine operators (cf. single-folded approach [18], [25]). In other words, an attacker has to guess only all necessary human operators for reconstructing an inner private key. In this scenario, the probability of bribing less than 5 assigned operators tends towards zero. If in average 10 operators, which corresponds to nearly 5 percent of all 211 operators, act corruptly, the probability of compromising the privacy of a certain user still only amounts to less than 0.1 percent. Regarding a number of 30 bribed operators, the probability raises up to slightly more than 2 percent. Hence, even the application of a single-folded threshold scheme provides appropriate security.

As aforementioned, adding machine operators to the system with concurrent application of our two-folded approach helps to balance the operational costs and the security level of the system. The second column of table III presents the probabilities of re-establishing the subsecret σ_M — 30 percent and 70 percent — which seems rather high, if taken out of the context that we applied a two-folded threshold scheme approach. However, reconsidering the above-mentioned constraints in our example implies, that a successful attack is still decreased by adding only a few machine operators. In other words, if there were solely human operators, the probability

of guessing at least all necessary shares for a certain user by bribing 20 human operators would amount to less than 0,7 percent, whereas applying our two-folded secret sharing scheme additionally reduces this probility to slightly less than 0,2 percent for 2 bribed machine operators respectively approximatly 0,5 percent for 3 bribed machine operators.

In the next section we provide the formal workflow of recovering a lost key.

B. Recovering a Lost Key

If a patient has lost her smart card, she identifies against a human operator. To rebuild a lost smart card it is not necessary that this human operator holds a part of the patient's inner private key. The human operator starts the recovering process by sending a message to the logic.

Necessary operations: proof patient's identity

2: $L \rightarrow O$:$\left\{\{A_{id}\}_{K_L}\right\} \forall\ \mathcal{O}$

A message is generated by the central logic and sent to all operators $\mathcal{O}$ with an encrypted version of the patient's identifier A_{id}. As mentioned in the previous section, to hide the patient's identifier A_{id} the logic key K_L is applied.

Necessary operations: encrypt patient's identifier

3: $O \rightarrow L \rightarrow St$:$\left\{\left\{\{A_{id}\}_{K_L}\right\}_{\overline{K}_O}\right\} \forall\ \mathcal{O}$

If an operator receives the message, she looks up her backup keystore via the central logic by encrypting the cipher text with her inner symmetric key $\overline{K}_O$. The central logic is able to find out if an operator possesses a patient's key share with this message.

Necessary operations: encrypt shares by $|\mathcal{O}|$ operators

4a: $St \rightarrow L \rightarrow M$:$\left\{\left\{\left\{\sigma_{\mathcal{M}_\iota}(\widehat{K}_A^{-1})\right\}_{K_L}\right\}_{\overline{K}_M}\right\} \forall\ \mathcal{M}$

4b: $St \rightarrow L \rightarrow H$:$\left\{\left\{\left\{\sigma_{\mathcal{H}_\iota}(\widehat{K}_A^{-1})\right\}_{K_L}\right\}_{\overline{K}_H}\right\} \forall\ \mathcal{H}$

After querying the double encrypted ciphertexts against the storage, the logic receives the associated double encrypted key shares and forwards them to the assigned human and machine operators.

Necessary operations: $|\mathcal{H}| + |\mathcal{M}|$ SQL select statements

5a: $H \rightarrow L$:$\left\{\left\{\sigma_{\mathcal{H}_\iota}(\widehat{K}_A^{-1})\right\}_{K_L}\right\} \forall\ \mathcal{H}^n$

5b: $M \rightarrow L$:$\left\{\left\{\sigma_{\mathcal{M}_\iota}(\widehat{K}_A^{-1})\right\}_{K_L}\right\} \forall\ \mathcal{M}^n$

The next step is that all assigned human operators decrypt their particular shared secrets with their inner symmetric key $\overline{K}_H$ and transmit them to the logic. The machine operators conduct the mirrored operation with their shared secrets and their inner symmetric keys $\overline{K}_M$. The logic is now able to decrypt these shares with its key K_L and consequently to combine the human shared secrets $\sigma_{\mathcal{H}_\iota}$ as well as the machine shared secrets $\sigma_{\mathcal{M}_\iota}$. As soon as the logic receives the shares from a minimum number of $k_H \wedge k_M$ necessary operators, the patient's inner private key can be re-calculated with appliance of Shamir's threshold scheme [33].

Necessary operations: decrypt a maximum of $|\mathcal{H}^n| + |\mathcal{M}^n|$ key shares, apply threshold scheme

6: $L \rightarrow St$:$\left\{\left\{\widehat{K}_A^{-1}\right\}_{K_{A'}}\right\}$

Next, the logic generates a new outer key pair $(K_{A'}, K_{A'}^{-1})$ which replaces the lost outer keys (K_A, K_A^{-1}) of the smart card. The logic encrypts the patient's inner private key with the new outer public key $K_{A'}$ and saves this ciphertext in the storage. Additionally, a new smart card is produced by the logic. Finally, the storage deletes the operator shares and their relations to the patient to avoid replay-attacks.

Necessary operations: generate new asymmetric key pair, encrypt patient's inner private key

7a: $L \rightarrow H$:$\left\{\left\{\left\{\sigma_{\mathcal{H}_\iota}(\widehat{K}_A^{-1}), A_{id}\right\}_{K_L}\right\}_{\widehat{K}_H}\right\} \forall\ \mathcal{H}^n$

7b: $L \rightarrow M$:$\left\{\left\{\left\{\sigma_{\mathcal{M}_\iota}(\widehat{K}_A^{-1}), A_{id}\right\}_{K_L}\right\}_{\widehat{K}_M}\right\} \forall\ \mathcal{M}^n$

Subsequently, the logic randomly chooses $\mathcal{H}^n \wedge \mathcal{M}^n$ assigned operators and uses the threshold scheme to divide the patient's inner private key into two shares, one for the human $\sigma_{\mathcal{H}}(\widehat{K}_A^{-1})$ and one for the machine operators $\sigma_{\mathcal{M}}(\widehat{K}_A^{-1})$. Once more, all shares will be double-enveloped. Firstly, the logic applies its key K_L and secondly, encrypts the gained ciphertexts with the certain inner public keys $\widehat{K}_H$ of the selected human or $\widehat{K}_M$ of the selected machine operators. These encrypted secret shares will then be transmitted to the operators. Moreover, the logic applies the same encryption procedures to the patient's id A_{id} and transfers this ciphertext to the operators, too.

Necessary operations: apply threshold scheme, encrypt shares and patient's identifier twice for $|\mathcal{H}^n| + |\mathcal{M}^n|$ operators

8a: $H \rightarrow L \rightarrow St$:$\left\{\left\{\left\{\sigma_{\mathcal{H}_\iota}(\widehat{K}_A^{-1}), A_{id}\right\}_{K_L}\right\}_{\overline{K}_H}\right\} \forall\ \mathcal{H}^n$

8b: $M \rightarrow L \rightarrow St$:$\left\{\left\{\left\{\sigma_{\mathcal{M}_\iota}(\widehat{K}_A^{-1}), A_{id}\right\}_{K_L}\right\}_{\overline{K}_M}\right\} \forall\ \mathcal{M}^n$

Upon receipt, the assigned human and machine operators decrypt their particular shares and the patient's identifier with their inner private keys $\widehat{K}_H^{-1}$ / $\widehat{K}_M^{-1}$. Then they again encrypt both attributes with their inner symmetric keys $\overline{K}_M$ / $\overline{K}_H$ and return these ciphertexts to the logic which saves them in the storage.

Necessary operations: decrypt and encrypt the key shares and the patient's identifier for $|\mathcal{H}^n| + |\mathcal{M}^n|$ operators; $|\mathcal{H}^n| + |\mathcal{M}^n|$ SQL insert statements to store the ciphertexts in the database

This workflow shows how the access to the backup keys and their re-establishing works. Moreover, with this technique we assure, that a patient's old smart card is not useable any more.

V. Conclusions

The introduction of the EHR promises massive savings [1]–[3] and a better service quality [4]–[7] for the patients. Moreover, as modern health care systems still lack standard processes, such a system could also support the definition and execution of e-Health workflows [8], [9]. Several approaches have been proposed to solve the challenge for implementing the EHR (cf. [13], [16], [17], [34]), but these architectures have vulnerabilities regarding their security. For example, current approaches (i) rely on a centralized patient-anamnesis list, which could be attacked form in- or outside and (ii) the dependency on a single pseudonym could lead to a data-mining or profiling attack because an attacker may guess the patient based on her medical history [18], [19], [26]. As medical data, for example a HIV-infection or the overall state of a patient's health, is sensible information, it is necessary to assure the patients' privacy.

Security is a balancing act between protection, performance and usability. Our system PIPE [18], [19], [25], [26], is based on a novel pseudonymization architecture. Our security hull model provides a patient with full control to her data by handing over the possibility of authorization by encryption, which currently is the most secure technique. To get access to the pseudonymized data in our system every participant possesses a smart card. As this security token may be lost, stolen, compromised or just worn-out, we proposed the idea of applying a threshold scheme to provide a secure backup. This approach significantly decreases the risk of attacks on the backup keystore.

In this publication we outlined the security of this technique and stated the necessary expenses to be able to recover access to the key and consequently to the pseudonymized data. We showed how human and machine operators can be set in to lower the costs of our fall-back mechanism. We outlined that our backup approach may still be regarded secure even if more then one tenth of all operators have been bribed. Finally, we stated the formal workflow of recovering a user's key.

The next steps in our research project are to deploy our prototype at our business partner's network and conduct tests in a medical environment.

Acknowledgment

We want to thank Karl Grill and Erich Neuwirth for their support on our statistical model and Stefan Jakoubi for his review.

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] S. J. Wang, B. Middleton, L. A. Prosser, C. G. Bardon, C. D. Spurr, P. J. Carchidi, A. F. Kittler, R. C. Goldszer, D. G. Fairchild, A. J. Sussman, G. J. Kuperman, and D. W. Bates, "A cost-benefit analysis of electronic medical records in primary care," *The American Journal of Medicine*, vol. 114, pp. 397–403, 2003.

[2] F. R. Ernst and A. J. Grizzle, "Drug-related morbidity and mortality: Updating the cost-of-illness model," University of Arizona, Tech. Rep., 1995.

[3] ——, "Drug-related morbidity and mortality: Updating the cost-of-illness model," University of Arizona, Tech. Rep., 2001.

[4] 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*, 2001, pp. 849–854.

[5] R. B. Elson and D. P. Connelly, "Computerized patient records in primary care. their role in mediating guideline-driven physician behavior change," *Family Medicine*, vol. 4, no. 8, 1995.

[6] M. H. Trivedi, J. K. Kern, B. D. Grannemann, K. Z. Altshuler, and P. Sunderajan, "A computerized clinical decision support system as a means of implementing depression guidelines," *Psychiatric Services*, vol. 55, pp. 879–885, 2004.

[7] D. W. Bates, "Medication errors : How common are they and what can be done to prevent them?" *Drug safety*, vol. 15, pp. 303–310, 1996.

[8] E. A. McGlynn, S. M. Asch, J. Adams, J. Keesey, J. Hicks, A. DeCristofaro, and E. A. Kerr, "The quality of health care delivered to adults in the united states," *The New England Journal of Medicine*, vol. 348, pp. 2635–2645, 2003.

[9] L. L. Leape and D. M. Berwick, "Five years after to err is human - what have we learned?" *Journal of the American Medical Association*, vol. 293, pp. 2384–2390, 2005.

[10] Health Level Seven, Inc., "Hl7 version 3," online: http://www.hl7.org/v3ballot/html/welcome/environment/index.htm, September 2007.

[11] IBM, "Machbarkeitsstudie betreffend Einfuehrung der elektronischen Gesundheitsakte (ELGA) im oesterreichischen Gesundheitswesen," 2006.

[12] R. 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.

[13] 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.

[14] G. de Moor, B. Claerhout, and F. de Meyer, "Privacy enhancing technologies: the key to secure communication and management of clinical and genomic data," *Methods of information in medicine*, vol. 42, pp. 148–153, 2003.

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

[16] K. Pommerening, "Medical Requirements for Data Protection," in *Proceedings of IFIP Congress, Vol. 2*, 1994, pp. 533–540. [Online]. Available: citeseer.ist.psu.edu/330589.html

[17] K. Pommerening and M. Reng, *Medical And Care Compunetics 1*. IOS Press, 2004, ch. Secondary use of the Electronic Health Record via pseudonymisation, pp. 441–446.

[18] B. Riedl, V. Grascher, and T. Neubauer, "Applying a threshold scheme to the pseudonymization of health data," in *to appear in the proceedings of the 13th IEEE Pacific Rim International Symposium on Dependable Computing (PRDC07)*, 2007.

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

[20] R. C. Barrows and P. D. Clayton, "Privacy, confidentiality, and electronic medical records," *Journal of the American Medical Informatics Association*, vol. 13, pp. 139–148, 1996, http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=116296&blobtype=pdf.

[21] K. Pommerening, "Pseudonyme ein Kompromiss zwischen Anonymisierung und Personenbezug," *Medizinische Forschung - Aerztliches Handeln; 40. Jahrestagung der GMDS*, pp. 329–333, September 1995.

[22] M. Ackerman, R. Craft, F. Ferrante, M. Kratz, S. Mandil, and H. Sapci, "Telemedicine technology," *Telemedicine Journal and e-Health*, vol. 8, No. 1, pp. 71 –78, 2002.

[23] J. Montagnat, F. Bellet, H. Benoit-Cattin, V. Breton, L. Brunie, H. Duque, Y. Legr, I. E. Magnin, L. Maigne, S. Miguet, J. M. Pierson, L. Seitz, and T. Tweed, "Medical images simulation, storage, and processing on the european datagrid testbed," *Journal of Grid Computing*, vol. 2, Number 4, pp. 387–400, 2004.

[24] A. Pfitzmann and M. Koehntopp., "Anonymity, Unlinkability, Unobservability, Pseudonymity, and Identity Management A Consolidated Proposal for Terminology," in *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2005.

[25] B. Riedl, T. Neubauer, and O. Boehm, "Patent: Datenverarbeitungssystem zur verarbeitung von objektdaten," *Austrian-Provisional-Application, Application No. A 1928/2006*, 2006.

[26] B. Riedl, T. Neubauer, G. Goluch, O. Boehm, G. Reinauer, and A. Krumboeck, "A secure architecture for the pseudonymization of medical data," in *Proceedings of the Second International Conference on Availability, Reliability and Security*, 2007, pp. 318–324.

[27] K. 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*, vol. 9, 2004.

[28] D. Lobach and D. Detmer, "Research challenges for electronic health records," *American Journal of Preventive Medicine*, vol. 32, Issue 5, pp. 104–111, 2007.

[29] T. Thornburgh, "Social engineering: the "Dark Art"," in *Proceedings of the 1st annual conference on Information security curriculum development*. New York, NY, USA: ACM Press, 2004, pp. 133–135.

[30] K. Maris, "The Human Factor," in *Proceedings of Hack.lu, Luxembourg*, 2005.

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

[32] J. Han and M. Kamber, *Data mining: concepts and techniques*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000.

[33] A. Shamir, "How to share a secret," *Commun. ACM*, vol. 22, no. 11, pp. 612–613, 1979.

[34] V. Schmidt, W. Striebel, H. Prihoda, M. Becker, and G. D. Lijzer, "Patent: Verfahren zum be- oder verarbeiten von daten," *German Patent, DE 199 25 910 A1*, 2001.

[35] D. C. Wherry, "Secure your public key infrastructure with hardware security modules," SANS Institute, Tech. Rep., 2003.

[36] Federal information processing standards publication, "Security requirements for cryptographic modules (fips pub 140-2)," Institute of Standards and Technology (NIST), Tech. Rep., 05 2001.

[37] Office for National Statistics (ONS), "T 04: England; estimated resident population by single year of age and sex; mid-2005 population estimates," ONLINE, 08 2006, http://www.statistics.gov.uk/statbase/Product.asp?vlnk=14508&More=Y.

[38] ——, "Average total income and average income tax payable: by sex, 2000/01: Regional trends 38," ONLINE, 2002, http://www.statistics.gov.uk/StatBase/ssdataset.asp?vlnk=7752. [Online]. Available: http://www.statistics.gov.uk/StatBase/ssdataset.asp?vlnk=7752

[39] J. Monger, "International comparisons of labour disputes in 2002," *Labour Market Trends*, vol. 111, pp. 19–28, 2003.