International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 5, Issue 4, July - August 2016 **ISSN 2278-6856**

Pseudonymization Techniques for Providing Privacy and Security in EHR

Bipin Kumar Rai[1], Dr. A. K. Srivastava[2]

[1]Research Scholar (Banasthali University, Rajasthan) & Associate Professor(CSE) ABESIT, UP, India

[2] Professor, CSE Department, Rayat Bahra RIMT, Sonepat, Haryana, India

Abstract: ***Pseudonymization is a technique where identification data is replaced by pseudonyms which are identifiers of subjects. Pseudonymization can be used for the privacy, confidentiality and integrity issues. Another scope of pseudonym is linkability. In this paper we analyzed different pseudonymization techniques for Electronic Health Record (EHR) and found more suitable for the healthcare information system.***

Key Terms: EHR, PHR, EMR, Pseudonymization, Anonymization.

1. INTRODUCTION

Security and privacy is primary concern in healthcare for online access of Electronic Health Record (EHR).This online accessing of patient record and transaction related to diagnosis have lots of benefits for patients as well as healthcare organization and professionals. But there are some serious privacy issues related to private data of patient e.g. any patient would not like to expose some sensitive health information which may cause problem for his professional career or defame him [1].

Internet based Electronic Health Record (EHR) systems allow to patient remote accessing to their entire medical information anytime. Electronic Health record (EHR) systems are highly desired for efficient integration of all relevant medical information of a person and to represents a lifelong documentation of the medical history. Several threats to confidentiality of healthcare information from inside the patient care institution or from within secondary user setting are crucial. Inside patient care institution consists of accidental disclosure, insider curiosity, insider subornation. Unauthorized access is required to carefully control. Outsider intrusion into medical information systems are required to control properly. [2]

Pseudonymization is a technique where identification data is replaced by pseudonyms which are identifiers of subjects. Pseudonym is a bit stream which is unique as identifier and suitable to authenticate the holder and his data. This pseudonym can only be associated with identification data by some secret value.

Pseudonymization can be used for handling the privacy issues, confidentiality as well as integrity. Another scope of pseudonym is linkability i.e. the knowledge of the relationship between the holder and his pseudonym. This information should be known to holder only (or trusted third party (TTP)).There are two ways to generate globally unique pseudonyms for holder. [3]

1. Centralized generation: A centralized third party may generate the pseudonym on the behalf of holder. Normally a hierarchical organized issuing party is used for large scale. The holder of the certificate has to trust on the issuer.

2. Holder-based generation: The holder generates his pseudonym locally. Only holder is aware about the relation between his identity and pseudonym. The holder locally generates globally unique random pseudonym. A mechanism is required to prove that holder generated a specific pseudonym without disclosing his identity and he can reveal his identity by disclosing the pseudonym.

2. PSEUDONYMIZATION APPROACHES

2.1 Peterson approach [4]

- In this approach a unique Global key (GK) and server side key (SSID) is provided when user get registered at service provider's website.

- Also a unique personal encryption key (PEK) and a password has to provide. An ID card is used which includes GK. User retrieves data from database by entering GK and PEK.
- It consists of three database tables---Data table, User table and security table (used to links data in data table to appropriate entry in user table).
- Data is encrypted two times before storing in the database.
- By knowing GK or PEK or both, anyone can view medical data without knowledge of password. (As data does not contain indentifying information).

For modification/add/delete of medical data, password is needed.

- It have a fall back mechanism for lost of GK. It requires PEK and password to get new GK.

- Whenever patient's data altered a secret password known only to patient is needed and hence unauthorized access of data is prevented.
- Patient can access their data via internet by accessing medical records database. Although data is encrypted but it is invisible to user.
- For emergency access of patient data, patient select second unique identifier other than GK. This is sufficient to access patient data without need of any password. This is the point where attack is possible.
- Patients have control over their data as they can modify any data including password, GK, PEK.
- PEK is easy to remember but complex enough to guess as it can be choose by using native language character. These characters are entered by using special input method.
- This is menu driven system for data input, simplifies task of translation.
- Ability of immediate change of access keys, lock-out any identity thieves.
- Privacy is achieved; summaries of medical records can also be accessed any time from internal terminal.

The major issues with this approach are—

1. All keys used for decryption of medical data are stored in the database. By accessing database, data may be changed as password and keys are stored in the database.
2. Unique PEK is selected by user—security leak --as it provides information about the existing keys.

2.2 Pseudonymization of Information for Privacy in e-health (PIPE)---

It is a solution, based on hull architecture instead of storing the relation between patients and their dataset in centralized manner [9].

- In this system all data are held persistently in the storage St, which consists of two separate databases. One hold plaintext pseudonyms and related medical datasets which are stored in plaintext for performance reasons. Other database is used to store user's personal information and their encrypted pseudonyms.
- The logic L is a centralized system for accessing St. This system is based on layered model, in which each layer comprises one or more secrets, like encrypted keys or hidden relations.
- To gain access to the secrets of one layer, any user will require secrets of the next outer layer.
- This model consists of three layers. The most inner layer consists of patients' diagnosis treatment and anamnesis datasets (ϕi). each of these entries is related to distinct pseudonym (ψij). These pseudonyms are shared with healthcare provider's to authorize them certain medical datasets.
- For full control over datasets a root pseudonym ψio for each ϕi is used. This is only known to patient and ensures that nobody except he is able to delete all pseudonym of certain datasets.

	Patient(A)	health care provider
Unique identifier	Aid	Cid
Outer (public key , private key)	(K_A , K'_A)	(K_C , K'_C)
Inner (public key , private key)	(iK_A , iK'_A)	(iK_C , iK'_C)
Inner symmetric key	SK_A	SK_C

- Pseudonym is encrypted in the next outer layer by SK_A. As SK_A is stored within the system, it is encrypted with iK_A. Furthermore iK'_A is encrypted with K_A.
- Only (K_A , K'_A) is available on smart card which is equipped with a logic chip to conduct encryption and decryption operations.
- For addition of a medical data by health care provider on behalf of patient both actor first of all authenticate against their particular smart card by entering a PIN. They use K'u to decrypt iK'u and subsequently SKu.
- The patient and healthcare provider authenticate against the system and establish a secure channel between them and the logic L. afterwards A and C send their Aid and Cid to L respectively. L sends Cid to A and Aid to C.

  now A→ L → St : E_{SKA}(Aid, Cid) and C→ L → St : E_{SKc}(Aid, Cid). Subsequently storage replies with the particular iK_A iK_c to L. Now logic sends to store and C. L → St: E_{iKA}(Cid, ψio, ψij) and L → C : E_{iKc}(Aid, ψij). Then C replies to L
  C→ L: E_{SKc}(Aid, ψij , Cid, tag) , ϕi
  For integrity purpose C sends following m-
  C→ L: E_{SKc}($fsign_{iK'c}$(fhash(ϕi))) || $fsign_{iK'c}$(fhash(ϕi))
- It implements a secure fall back mechanism when smart card is lost.

2.3 Electronic health card (eGK)—

- It is a designed as service –oriented architecture (SOA) having some restrictions like local card access only, RMI communication, supported by ministry of health Germany.
- It provides application like e-prescription, EMR, EHR, emergency data etc. its architecture consists of five layers.

A. Presentation layer--to provide communication interface to user

B. Service layer—to provide different services

C. Business layer---to combine different services

D. Application layer-manages data and user right

E. Infrastructure layer

- To authenticate a user, system encrypts a random number with public key of user and hence user has to decrypt with his private key which is stored in his card.
- Data is encrypted with a one-time symmetric key i.e. session key. This session key is encrypted with the public key of the patient which is decrypted by his private key. Finally the data is decrypted with this session key.
- A file has a default ticket toolkit and any number of private ticket toolkit(user defines for other user). This kit consists of a ticket building tool, a ticket verifier, list of access policy and encrypted link to file.
- eGK optionally provide permission to store private ticket toolkit for every entry which uses asymmetric key pair stored on emergency card. In case of card lost, this emergency card is used to decrypt the session keys of second ticket toolkit. At last, with the keys of new card the session keys are encrypted.
- eGK store data in the database after pseudonymization. So database access does not mean linking of identification with the data.
- Strong fall back mechanism

2.4 Thielscher Approach [5]

- Identification data and the anamnesis data of medical record are stored in two different databases. For the retrieval of health data requires data record identifier code assigned to each patient. This code is detached from patient identifying data. It includes a patient card code stored on patient card and patient identification code(PIN).
- Decentralized keys stored on smart cards which is used to link the patient identity to his data. To generate a unique data identification code (DIC), which is also stored in the database this key is used.
- DIC is totally independent to identifying data. It is shared healthcare entity and patient for limited period.
- When health data is requested by a profession like physician then electronic patient card and an identification code of that professional is required. So in this way system can have the usage history of the patient data.
- Transfer of DIC or patient data from centralized database is accessed in encrypted mode. Hence unauthorized interception of data record of identifier code is protected.
- Electronic health card contains picture of the patient to identifying him. The health professional can match at time of treatment.
- Pseudonymization computer physically separate from centralized database without any on-line connection, is used. It replaces person identifying data with the corresponding data record identifier code. Now this updated data is provided for online- access.
- Some part of stored health data in database is also stored on electronic card so that health professional can know the status of patient without directly.
- A fall back mechanism is provided.
- Offline storage of pseudonym hold by patient. It is the shortcoming of this approach that centralized pseudonym is centrally stored in the patient mapping list for recovery purposes which is open to insider abuse.

2.5 Pommerening Approach [6]

- Different approaches for secondary use of medical data are proposed.
- In the first approach secondary user access and merge the data of patient but can not identify. It is based on data from overlapping sources for one-time secondary use.
- A unique identifier (PID) is used to connect the data. A pseudonymization service encrypts the PID with a hash algorithm. With the public key of the secondary user the medical data is encrypted so secondary user can decrypt the medical data.
- The second approach is same as first but with the possibility to re-identify the patient. It stores a reference list of the patient's identity and the associated PIDs.
- The pseudonymization service decrypts the pseudonym and sends the request to the PID service, which permits notifying the data owner.

- The third approach is for many secondary users. A physician exports his local database to the central researcher Database.

- The identification data is replaced by a PID. Pseudonymization service is used for export of each secondary use of the data.

- The PID is encrypted by the by a project specific key to ensure that different projects get different pseudonyms.

2.6 Slamanig and Stingl Approach [7]

- This approach considers two sets. One is set of users U= {U_1, U_2…..U_n } , U is a public user repository . Another is set of data objects represented as documents D = {D_1, D_{2s} ,……..D_m }, and D is document repository. Both repositories can be located in different locations.

- Centralized component which serves as a point of access to the set of documents. Document repository maps these references to respective documents.

- Stores the data in a centralized database and uses smart cards for authentication.

- A E-health portal implements an authorization concept which can be described by means of a relation R defined over $U^4 \times D$ where every 5-tuple (Us, Ur, Uc , Up, Dj) $1 \leq s,r,c,p \leq n$ and $1 \leq j \leq m$ represents an access right for a document Dj user Us (sender) grants user Ur (receiver) created by Uc concerning patient Up.

- To ensuring unlinkability it uses a combination of pseudonym and anonymous authentication. The ehealth portal can be classified by means of personalization of their offered services.

- The system keeps the pseudonyms of a user secret. Each pseudonym realizes a sub-identity of the user and is encrypted with a public key.

- To access datasets of one of his sub-identities, user has to login into the system with his general pin code. Also he has to enter the pin code of the sub-identity to activate the private key on the smart card.

- Pseudonymized portals- it can be described by means of encryption of the content data and the additional protection of the metadata of the system. Metadata is protected via a mechanism, pseudonymization. Every user Ui choose randomly a second identifier P_{Ui} i.e. pseudonym which is used to identify his share.

- To prevent linkage between user and his pseudonym, it is stored in encrypted form $E_{Ui}(P_{Ui})$ in the user repository. The unlinkability holds since P_{Ui} is independently chosen. In order to hide links between users and documents every elements of the tuple need to be pseudonymized

  $t_{Us \to Us} = (E_{Us} (Us), P_{Us} , E_{Us} (Uc), E_{Us} (Up), E_{Us} (Dj , k_{Dj}))$
  Here sender, receiver and creator are one, thus this share represents an access right which is given from the creator of a document to himself. In case of a share granted by user Us to user Ur for the document Dj every access right can be represented by pair of 5-tuple ($t_{Us \to Us,} t_{Us \to UR}$).
  $t_{Us \to Us} = (E_{Us} (Us), P_{Us} , E_{Us} (Uc), E_{Us} (Up), E_{Us} (Dj , k_{Dj})$
  $t_{Us \to UR} = (E_{UR} (Us), U_R , E_{UR} (Uc), E_{UR} (Up), E_{UR} (Dj , k_{Dj}))$
  Any other user is able to identify Ur as receiver of this share. Therefore Ur needs to modify the share and replace $t_{Us \to UR}$ with $t'_{Us \to UR}$.
  $t'_{Us \to UR} = (E_{UR} (Us), P_{UR} , E_{UR} (Uc), E_{UR} (Up), E_{UR} (Dj , k_{Dj}))$
  Since P_{UR} represents a pseudonym of user Ur and is solely known to him, the unlinkability between the user and the share is guaranteed.

- Distributed key backup to *N* users using a (*t*,*N*)-threshold secret sharing scheme suggested as fall-back mechanism.

3. CONCLUSION

Pseudonymiztion techniques are suitable technique for EHRs. It should be properly designed to overcome different privacy issues. Pseudonimization based patient controlled EHR system may be well accepted solution which can handle all the security and privacy issues.

REFERENCES

[1.] Rui Zhang, Ling Liu,"security models and requirements for healthcare application clouds", IEEE 3rd international conference on cloud computing, 2010.

[2.] Rima Addas, Ning Zhang," Support access to distributed EPRs with three levels of identity privacy preservation", sixth international conference on availability, reliability, and security, IEEE computer society 2011.

[3.] Scharter, Schaffer, " Unique user-generated digital Pseudonyms", springer LNCS 3685 september 2007.

[4.] Peterson, R.L.: Encryption system for allowing immediate universal access to medical records while maintaining complete patient control over privacy. US Patent Application Publication, No.: US 2003/0074564 A1 (2003).
[5.] Thielscher, C., Gottfried,M., Umbreit, S., Boegner, F., Haack, J., Schroeders, N.: Patent:Data processing system for patient data. Int. Patent, WO 03/034294 A2 (2005)
[6.] Pommerening, K., Reng, M.: Secondary use of the Electronic Health Record via pseudonymisation. In: Medical And Care Compunetics 1, pp. 441–446. IOS Press, Amsterdam (2004).
[7.] Deniel slamanig,, Christian stingl , "privacy aspect of e-health" the 3rd international conference on availability, reliability and security, IEEE computer society 2008.
[8.] Bernhard riedl, Veronica ," Assuring Integrity and confidentiality for pseudonymized health data", IEEE 3rd international conference on Availability, Reliability , and security, 2008.
[9.] Benhard Riedl, Grascher, Fenz , Neubauer , "Pseudonymization for Improving the privacy in e-health applications ", IEEE 41st Hawaii International conference on system sciences, 2008.

AUTHOR

Bipin Kumar Rai received the B.Tech (CSE) from UPTU(BIT Muzaffarnagar) Lucknow, UP and M.Tech (CSE) from RGPV Bhopal, (SSSIST, Sehore) MP in 2004 and 2009, respectively. During 2004-2006&2008 to 2014, he taught in different engineering colleges. He is with ABESIT as Associate Professor now. His area of interest is Information Security.

Prof. Anoop Kumar Srivastva is a potential researcher and scientist in the domain of Artificial Intelligence and information security. He received his B.Tech.(EE) from KNIT Sultanpur,India and PhD(AI) from TIFR, Mumbai, India. He taught in different engineering colleges as Director/Prof. Currently he is with Rayat Bahra RIMT, Sonipat (Haryana) as Director.