Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Privacy preservation and information security protection for patients' portable electronic health records

Lu-Chou Huang[a,b], Huei-Chung Chu[b], Chung-Yueh Lien[a], Chia-Hung Hsiao[c], Tsair Kao[d,*]

[a] *Institute of Biomedical Engineering, National Yang-Ming University, Taipei, Taiwan*
[b] *Department of Information Management, Hua-Fan University, Taipei, Taiwan*
[c] *Department of Medical Informatics, Tzu Chi University, Hualien City, Taiwan*
[d] *Department of Biomedical Engineering, Hungkuang University, Shalu, Taichung County, Taiwan*

ARTICLE INFO

Article history:
Received 19 December 2007
Accepted 15 June 2009

Keywords:
Privacy
Security
De-identification
Portable electronic health record

ABSTRACT

As patients face the possibility of copying and keeping their electronic health records (EHRs) through portable storage media, they will encounter new risks to the protection of their private information. In this study, we propose a method to preserve the privacy and security of patients' portable medical records in portable storage media to avoid any inappropriate or unintentional disclosure. Following HIPAA guidelines, the method is designed to protect, recover and verify patient's identifiers in portable EHRs. The results of this study show that our methods are effective in ensuring both information security and privacy preservation for patients through portable storage medium.

1. Introduction

In 1996, the Health Insurance Portability and Accountability Act (HIPAA) offered some general guidelines to enforce the protection of private medical information. One such guideline stated that patients must be able to view and obtain copies of their records, and request amendments to confirm they have the right of accessing their medical records to understand and monitor their health status and the process of diagnosis and therapy [1–3]. In the real world, patients' health records are distributed around different hospitals and clinics, and the retrieval of this scattered information when a patient visits a doctor in any particular hospital is a major problem. Currently, there were two ways to overcome this problem: either the patient can carry his/her own records, or the records can be transmitted through an electronic network. For example, consider an emergency circumstance where a patient is seen by a doctor in a different hospital than the one he/she normally goes to. If the hospitals are already in collaboration for sharing electronic medical records, the doctor can get the patient's history through the Internet. This type of exchange typically offers more extensive auxiliary reference materials to doctors, and can therefore result in getting the best care. Nevertheless, if there is no existing collaboration, then the patient's history will be unavailable. Some hospitals and clinics are able to offer a patient's health history summary to the patient for taking to other hospitals for doctors' reference.

However, when patients assume control of their electronic health records (EHRs), there will inevitably be threats to jeopardize the security and privacy of their information. The responsibility for preserving the medical record shifts to the patient because the record is no longer under the hospital's protection. This could result in a violation of personal privacy due to the risk that the patient's medical records may become lost, stolen, disclosed, or distorted [4,5]. Some protective measures can help patients safeguard and store their medical records with portable data storage media (CDs/DVDs, diskettes, flash drives, etc.) after exiting the hospital. These measures aim to reduce or avoid any violations of personal privacy while providing increased opportunities for patients to seek consultation with other physicians or opinions from experts. From the patient's perspective, EHRs that offer portable use should:

- Have strong measures protecting confidentiality of the medical information they contain. [6–8] (*Requirement* I)
- Prove the validity and accuracy of the EHR so as to be able to protect the patient'S rights. [7] (*Requirement* II)
- Contain measures for the selective protection of privacy that allow for consultations with a trusted third party (TTP) on related medical information inquiries. [9–12] (*Requirement* III)

In *Requirements* I and II, the related patient's identifiable information is confidential, and security protection provides a way to guard that information. The solution is a method that applies cryptography

* Corresponding author.
E-mail address: tskao@sunrise.hk.edu.tw (T. Kao).

0010-4825/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiomed.2009.06.004

to archive security goals to protect the EHRs, such as the application of digital signatures, encryption algorithms and digital certificates [13,14]. In *Requirement* III, people desire anonymity for secondary use of health data, therefore the data must be de-identified or pseudonymization for privacy protection [15]. In order to achieve the purposes described above, the selective disclosure of personal information is allowed by patient control in conjunction with the patient's privacy policy [16–18]. Safeguarding a patient's identifiers is an essential step to protect his/her privacy [19–21]. Methods for de-identification differ based on the demands of different environments [22–29]. Some researchers have utilized the pseudonymization model for research purposes to encrypt patient identifiers in order to make patients' EHRs available for secondary use [30,31]. From the patients' perspective, if those advantages could be applied to EHRs with security protections for the recovery and duplication of records from the hospital, then it would be a useful change in the encryption of all EHR contents.

The aim of this study is to develop a software tool to identify different data types, such as text and image fields of a portable HER, to preserve the privacy and security to avoid any inappropriate or unintentional use or disclosure.

2. Methods

In this study, the EHRs we deal with include a summary of a patient's admission/discharge history and medical images from Changhwa Christian Hospital, Changhwa, Taiwan. We classify information in the EHR as being of the well-defined type or non-defined type. The well-defined type is a fixed structure form, such as Extensible Markup Language format (XML-based type) and Digital Imaging and Communication in Medicine (DICOM) file. The non-defined type, such as free text and non-DICOM format, has non-interoperability. In the image formats, the purpose of processing the image data is to deal with any identifiable text within the image. Our method divides an EHR into personally identifiable information (PII) and non-PII. The requirements and the applied technology for portable EHRs are shown in Table 1.

2.1. Proposed framework overview

The proposed method including the secure process and recovery process is illustrated in Fig. 1. The secure process includes de-identification, pseudonymity, patient selection, signing process, and encryption process. The recovery process includes re-identification and verification. After the signing and privacy preservation processes, there will be one file containing encrypted PII, one containing non-PII EHR, and two signatures. In order to verify the integrity of EHR, all of the encrypted PII must be decrypted and filled into the non-PII EHR to recover the original EHR. Encrypting all of the information is a special case in our approach, which regards all information as PII. The detailed descriptions of the method are as follows.

Table 1
The requirements for portable EHR.

Requirement		Implementation
Privacy protection	Protect identifiers	De-identification (HIPAA), Pseudonymity To find, extract and replace identifiers (automatically)
	Patient control	GUI for patient control
Security protection (**data**)	Confidentiality	Encryption/decryption
	Integrity	Digital signature
	Authentication	Digital signature
	Authorization	Smartcard/password
	Availability	Based on: 1. Authorized people (patient) 2. Authentication (file content)
	Non-repudiation	Digital signature Other support resources, such as TTP

HIPAA: Health Insurance Portability and Accountability Act; GUI: graphic user interface; TTP: trusted third party.

2.2. De-identification and pseudonymity

PII should be found by following HIPAA guideline before strengthening the protection measures. In the well-defined type of information, it is easy to find the PII from certain structures such as the XML schema patterns. In the non-defined type, the method utilizes the regular expressions, keyword filters, pre-defined area filters and optical character recognition (OCR) to recognize identifiers. The methods used in de-identification is shown in Table 2.

2.2.1. Regular expression method

This process utilizes regular expressions [27,32,33] to recognize identifiers through many formulations to find compatible patterns of information. Essentially, a regular expression is a string that describes a set of strings according to certain syntax rules. The process of regular expression rules includes two categories: (1) numerical data and (2) numerical data with specific characters and symbols. In category (1), several regular expression rules are used to perform pattern-matching to remove numerical identifiers. In category (2), expression rules with specific characters and symbols are used to detect the identifier patterns combined with vehicle number, national identification card number, medical record number, bed number, IP address, etc. Pre-defined regular expressions are stored in a template file. For example, Table 3 shows the different patterns of date with related regular expression rules.

2.2.2. Keyword filters

Use keyword filters to match patterns to recognize the identifiers and filter the context of matching keywords, which means the kind of tokens that appear before or after whole keyword phrases. For example, "Mr." is a trigger for name keyword with an offset of +2 as the usual pattern in text is "Mr. [name]". Another example is address in Taiwan, which can be represented as a sentence like "Section 2, Linong Street". The keywords "Section" and "Street" are recognized, and the phrases between "Section" and "Street" are filtered by keyword offsetting. We also defined template files that contain street names, geographic information, rare diseases and hospital information to filter identifiable information. We compiled the proper keyword look-up tables from three sources:

(1) The United National Mandarin Phonetic System of ChungHwa Post, which collects all of geographical locations in Taiwan such as building, alley, lane, street, road, neighborhood, village, district, township, city, county, province, and zip code.
(2) The Hundred Family Surnames, which is a classic Chinese text composed of common surnames in Chinese for identifying the names.
(3) Dictionary: we collected the various keywords from the Internet, including healthcare institute names, names of rare diseases, hospital basic information, department, facility name, etc.

2.2.3. OCR and pre-defined area filter for image data

Recognizing identifiers in the medical image, the method uses optical character recognition and pre-defined area filter. Applying OCR in the image is for the text of identifiers burned into the image pixel data. After OCR, it can get the text words and the positions of words in image, and then replace the pixels in the image from their related positions. Due to background noise during the processing of text in medical images with OCR, we did not intend to develop the

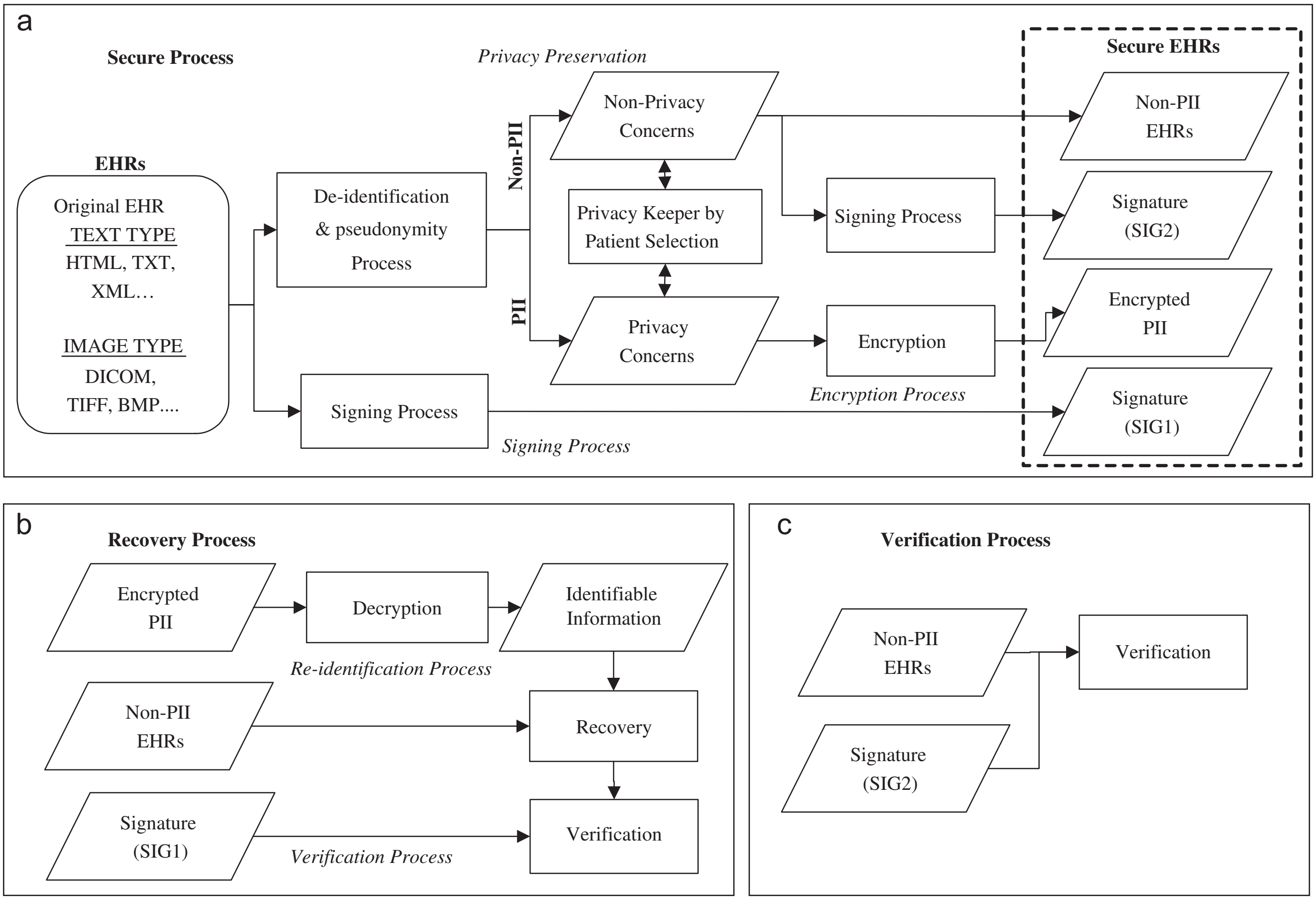


Fig. 1. Architecture of the system, (a) secure process, (b) recovery process and (c) verification process. There are two processes in the entire system: secure process and recovery process. It could deal with the text type and image type information within EHR. EHR in the text type include the well-defined type and non-defined type. In the secure process, signing process, privacy preservation process, privacy keeper by patient selection and encryption process are performed to provide a secure EHR for patients. In the recovery process, re-identification and verification functions are performed to obtain the original EHR. The supporting resources, such as certificate authority server and trusted third parties, could enhance the security.

Table 2
The method in de-identification process.

Identifiers (HIPAA)	Recognize identifiers method				
	Regular expression	Defined form	Defined keyword filter	OCR process	Other
1. Name	■	■			
2. Geographic information smaller than a state (i.e. city, zip code)	■	■	■		
3. Elements of dates including birth date, admission date, date of death, and all ages ≥ years of age	■				
4. Telephone numbers	■				
5. Fax numbers	■				
6. Electronic mail address	■				
7. Social security number	■				
8. Medical record number	■				
9. Health plan beneficiary numbers	■				
10. Account numbers	■				
11. Certificate of license numbers	■				
12. Vehicle identifiers and serial numbers including license plate	■				
13. Device identifiers and serial numbers	■	■			
14. Web universal resource locators (URLs)	■				
15. Internet Protocol (IP) address numbers	■				
16. Biometric identifiers, including finger and voice prints					■[a]
17. Full face photograph images and comparable images				■	■[a]
18. Any other unique identifying number, characteristics, or code characteristics, or code	■	■	■		

OCR: optical character recognition.

[a] Biometric identifiers and face photograph are not processed.

Table 3
Common regular patterns of date with related regular expression rules.

<table>
<tr><th>Common pattern (example date)</th><th>Related regular expression rules</th></tr>
<tr><td>yyyy/mm/dd: 2008/11/10,08/11/10</td><td>(?⟨Year⟩(? : \d{4}|\d{1,2}))/(?⟨Month⟩\d{1,2})/(?⟨Day⟩\d{1,2})(?x)</td></tr>
<tr><td>dd/mm/yyyy, 11/02/2008</td><td>(0[1-9]|[12][0-9]|3[01])[-/.](0[1-9]|1[012])[-/.](19|20)[0-9]{2}</td></tr>
<tr><td>mm/dd/yyyy, 02/11/2008</td><td>(0[1-9]|1[012])[-/.](0[1-9]|[12][0-9]|3[01])[-/.](19|20)[0-9]{2}</td></tr>
<tr><td>yy-m-d or yyyy-mm-dd: 1999-01-12, 1999.01.12</td><td>\b(19|20)?[0-9]{2}[-/.](0?[1-9]|1[012])[-/.](0?[1-9]|[12][0-9]|3[01])\b</td></tr>
<tr><td>yyyy-mm-dd, 2008-02-12</td><td>(19|20)[0-9]{2}[-/.](0[1-9]|1[012])[-/.](0[1-9]|[12][0-9]|3[01])</td></tr>
<tr><td>h:m:s</td><td>(\d{2}|\d{1}) : (\d{2}|\d{1}) : (\d{2}|\d{1})</td></tr>
<tr><td>yymmdd: 940101</td><td>9[01234]\d{4}</td></tr>
</table>

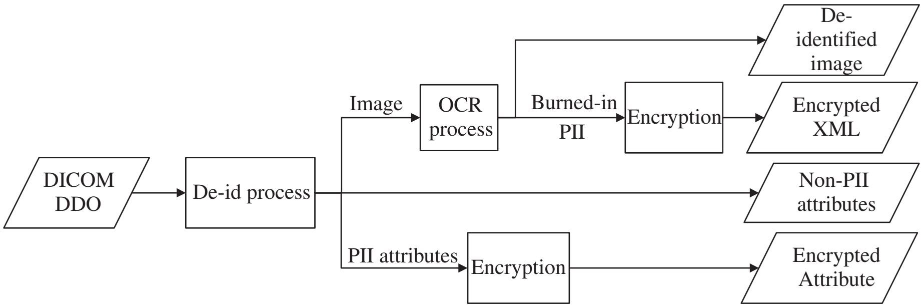


Fig. 2. DICOM image de-identification process.

OCR technique but merely to apply the OCR programs of Microsoft Office in our method. Therefore, we only focused on increasing accuracy by strengthening image pre-processing before its input to the system. In this study, the method can only recognize typewritten text in a medical image, but not hand-written print or cursive text.

2.2.4. DICOM de-identification process

Fig. 2 shows the de-identification process for a DICOM image. According to DICOM part 15 Security Profile [34], the PII and non-PII attributes are separated from a DICOM data object (DDO). The PII attributes are encrypted as Encrypted Attribute (0400, 0500). The burned-in PII will be removed from image by OCR process and be encrypted as the encrypted XML. The resultant image could be considered as a de-identified DICOM image.

2.3. Privacy keeper by patient selection

The International Federation of Gynecology and Obstetrics (FIGO) Committee stated that, “Patients have the right to ultimate control over the confidentiality of their data” [9]. Patient control allows customization to fit the patient’s individual needs [10–13]. After completing the automatic de-identification and pseudonymity processes, patients can use their own judgment to adjust the replaced phrase and image areas pertaining to their privacy and confidentiality. The method allows patients to select and review the process as well as preview expected results between the non-privacy concerns and the privacy concerns. They cannot modify original EHR content, but can control text and images that are related to the protected privacy issues. This provides a high level of personal control that a patient can exercise over their data.

2.4. Signing process

In Fig. 1(a), the original EHR is digitally signed (SIG1). The signature SIG1 is provided to verify the secure EHR after recovery process. In order to protect the integrity of non-PII, which is also digitally signed (SIG2). The signature SIG2 can be provided to verify the non-PII in the secure EHR. Signature SIG2 would be optional depending on its use. In order to verify the integrity of the original EHR, all of the PII must be decrypted and filled into the non-PII to recover the original EHR, as shown in Fig. 1(b). Moreover, the integrity of the non-PII can also be validated by SIG2, as shown in Fig. 1(c).

2.5. Encryption

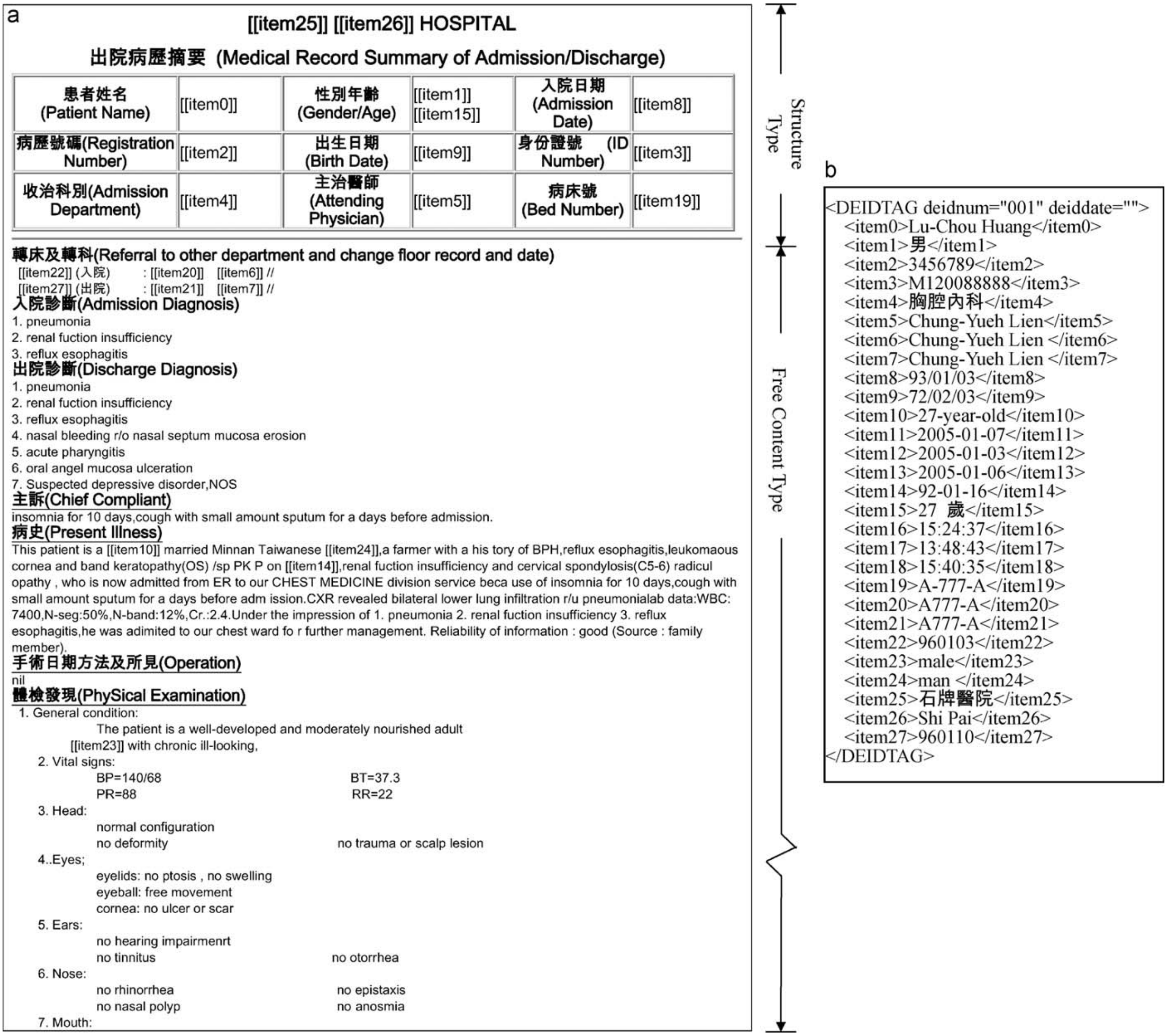
The method uses an encryption algorithm to encrypt PII after patient control. Only patients can decrypt their own PII with an encryption key. The encryption key is encrypted by password or using key transport mechanism. Key transport means that the encryption key has been encrypted with the patient’s public key using a public key algorithm. A personal identification number (PIN) to protect the patient’s private key stored in a smartcard can implement the key transport.

2.6. Re-identification and verification process

The re-identification process is shown in Fig. 1(b). The patient decrypts the encrypted PII by password or key transport. The decrypted PII needs to be combined with the non-PII to recover the original EHR. In key transport, the patient enters the PIN to access his/her private key from the smartcard and uses it to decrypt the encryption key. If the PIN is correct, the encryption key decrypts the cipher text to recover the encrypted PII. Using the signature SIG1 to verify recovered EHR, if the verification is valid, the recovered HER is the same as the original one. The signature SIG2 can verify the non-PII EHR in secure EHR, as shown in Fig. 1(c).

3. Results

The system operates under the Microsoft.NET Framework, and uses Microsoft Visual *C#* programming language. The OCR process utilizes Microsoft Office Document Imaging 2003 (MODI) as its library, and in the cryptographic process, we use a smartcard to input the key for the encryption and decryption work, which allows the system to run both symmetrical and asymmetrical algorithms. During the operating process, signatures must be confirmed



[[item25]] [[item26]] HOSPITAL

出院病歷摘要 (Medical Record Summary of Admission/Discharge)

患者姓名 (Patient Name)	[[item0]]	性別年齡 (Gender/Age)	[[item1]] [[item15]]	入院日期 (Admission Date)	[[item8]]
病歷號碼(Registration Number)	[[item2]]	出生日期 (Birth Date)	[[item9]]	身份證號 (ID Number)	[[item3]]
收治科別(Admission Department)	[[item4]]	主治醫師 (Attending Physician)	[[item5]]	病床號 (Bed Number)	[[item19]]

轉床及轉科(Referral to other department and change floor record and date)

[[item22]] (入院) : [[item20]] [[item6]] //
[[item27]] (出院) : [[item21]] [[item7]] //

入院診斷(Admission Diagnosis)

1. pneumonia
2. renal fuction insufficiency
3. reflux esophagitis

出院診斷(Discharge Diagnosis)

1. pneumonia
2. renal fuction insufficiency
3. reflux esophagitis
4. nasal bleeding r/o nasal septum mucosa erosion
5. acute pharyngitis
6. oral angel mucosa ulceration
7. Suspected depressive disorder,NOS

主訴(Chief Compliant)

insomnia for 10 days,cough with small amount sputum for a days before admission.

病史(Present Illness)

This patient is a [[item10]] married Minnan Taiwanese [[item24]],a farmer with a his tory of BPH,reflux esophagitis,leukomaous cornea and band keratopathy(OS) /sp PK P on [[item14]],renal fuction insufficiency and cervical spondylosis(C5-6) radicul opathy , who is now admitted from ER to our CHEST MEDICINE division service beca use of insomnia for 10 days,cough with small amount sputum for a days before adm ission.CXR revealed bilateral lower lung infiltration r/u pneumonialab data:WBC: 7400,N-seg:50%,N-band:12%,Cr.:2.4.Under the impression of 1. pneumonia 2. renal fuction insufficiency 3. reflux esophagitis,he was adimited to our chest ward fo r further management. Reliability of information : good (Source : family member).

手術日期方法及所見(Operation)

nil

體檢發現(PhySical Examination)

1. General condition:
    The patient is a well-developed and moderately nourished adult
    [[item23]] with chronic ill-looking,
2. Vital signs:
    BP=140/68 BT=37.3
    PR=88 RR=22
3. Head:
    normal configuration
    no deformity no trauma or scalp lesion
4..Eyes;
    eyelids: no ptosis , no swelling
    eyeball: free movement
    cornea: no ulcer or scar
5. Ears:
    no hearing impairmenrt
    no tinnitus no otorrhea
6. Nose:
    no rhinorrhea no epistaxis
    no nasal polyp no anosmia
7. Mouth:

```
<DEIDTAG deidnum="001" deiddate="">
   <item0>Lu-Chou Huang</item0>
   <item1>男</item1>
   <item2>3456789</item2>
   <item3>M120088888</item3>
   <item4>胸腔內科</item4>
   <item5>Chung-Yueh Lien</item5>
   <item6>Chung-Yueh Lien </item6>
   <item7>Chung-Yueh Lien </item7>
   <item8>93/01/03</item8>
   <item9>72/02/03</item9>
   <item10>27-year-old</item10>
   <item11>2005-01-07</item11>
   <item12>2005-01-03</item12>
   <item13>2005-01-06</item13>
   <item14>92-01-16</item14>
   <item15>27 歲</item15>
   <item16>15:24:37</item16>
   <item17>13:48:43</item17>
   <item18>15:40:35</item18>
   <item19>A-777-A</item19>
   <item20>A777-A</item20>
   <item21>A777-A</item21>
   <item22>960103</item22>
   <item23>male</item23>
   <item24>man </item24>
   <item25>石牌醫院</item25>
   <item26>Shi Pai</item26>
   <item27>960110</item27>
</DEIDTAG>
```

Fig. 3. A sample secure EHR: (a) the original EHR (~300 lines) after the de-identification and pseudonymity process. All identifiers are replaced with item tags. The original format is unchanged. (b) Extracted identifiable information and stored in XML format before encryption.

before the next step proceeds, and the private key must enter the PIN code.

The testing files used in this study were discharge summaries from a hospital. The algorithm failed when there was typographic or spelling error. The current version of the algorithm is tuned to patterns observed in discharge summaries, the approach may be customized to work on other free-text medical records. Our data held approximately 200,000 words over a sample volume of 200 admission/discharge summaries in HTML-based type. After the automatic de-identification and pseudonymity processes, the identifiers in the text portion of the EHR can be replaced with items completely. By applying both pre-defined templates and the OCR method, the identifiers in the image portion are replaced with black blocks. Fig. 3(a) shows how the method was used to replace identifiable text type information in the original EHR with tags. The identifiers in the XML before encryption are shown in Fig. 3(b).

The average percentage of PII was 4% in a record. Evaluation result of the process in recognizing identifiers the precision of each record was from 96.65% to 100%, and the average precision was 99.97%. The recall of each record was from 78.69% to 100%, and the average recall was 98.92%. F-measure was 88.07–100.00%, and the average was 99.41%. The algorithm failed when there was typographic or spelling error. The current version of the algorithm is tuned to patterns observed in discharge summaries, the approach may be customized to work on other free-text medical records.

The original image is shown in Fig. 4(a), and the same image after undergoing the de-identification process and having the identifiable area replaced with black pixels is shown in Fig. 4(b). When only the OCR method for recognition and processing was applied, some areas were not processed successfully. For example, in Fig. 4(a), the words "Changhwa Christian Hospital" are in the black and white border of the image, which will likely result in errors. The success rate of OCR was 60–100% in the 150 test images, and the averaged success rate was 65%. Despite recognizing almost all of the HIPAA identifiers during OCR process, some identifiers were missed due to various reasons. First, the font types and ambiguous characters such as "g",

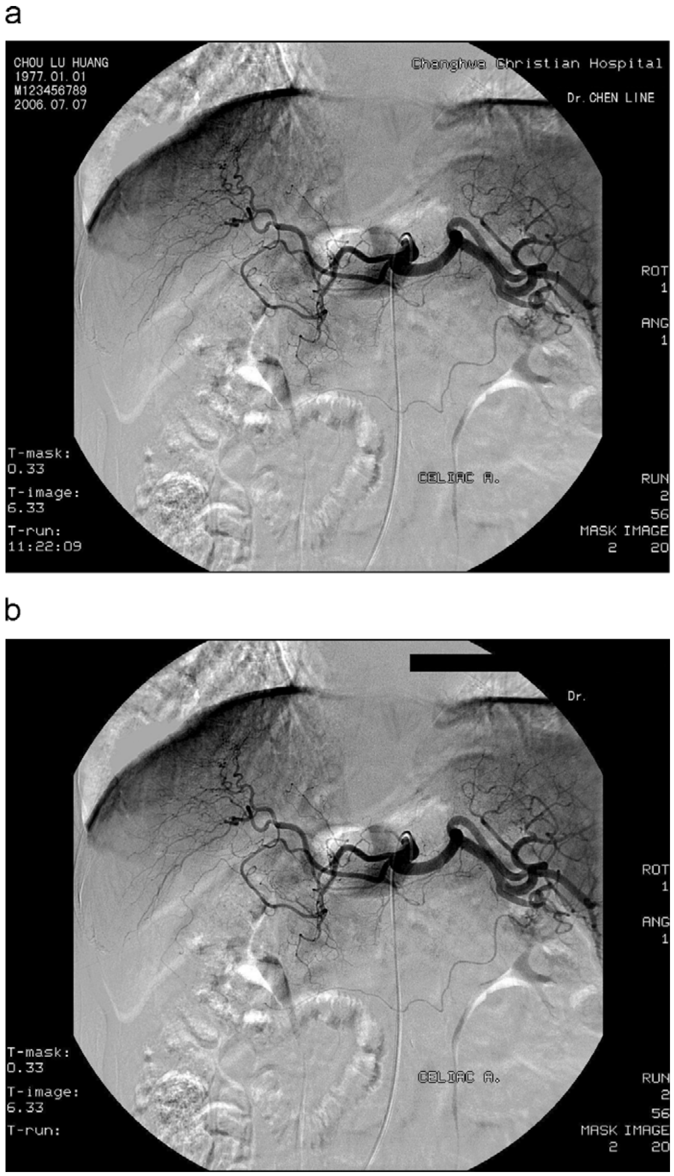


Fig. 4. The secure EHRs in image: (a) the original image before secure processing; (b) de-identified image of the original image. The identifiers in image are replaced with black pixels.

"9", "p", and "1", "l", "I", and "0", "o" are too similar to recognize; second, if the grey pixel value of background color is closer to that of the identifier; finally, if the identifiers are overlaid on the boundary edge of the images such as those of endoscope and ultrasound.

Using OCR on general documents with clear background has a low error rate. However, because the background of medical images is much more complicated, more errors could occur and further refinement is needed. Once the information was processed, patients could select the words and image regions that they are interested or concerned about, and correct any mistake made in the automatic process of de-identification and pseudonymity.

4. Discussion

With the public's increased use of the Internet, communication via e-mail has become widespread. There are some hospitals, physicians, or companies providing the services to allow patients upload their health records to seek second consultation for a clarification by e-mail or web service. These e-mails vary in degree of detail and specificity and almost through an essentially anonymous forum [35]. If patients do not want to expose their private information when they seek to obtain expert information, pseudonymous or de-identified EHR may be a good option.

The process of de-identification and pseudonym process are automatic to reduce the risk of damage in privacy by following HIPAA. In the well-defined type of EHR, the recognition of identifiers can be controlled completely. In the non-defined type, there are many ways to approach the recognition of identifiers. If the de-identification and pseudonymization results in incorrect or improper information, it is important for the patient to be able to correct it. Even though the algorithm correctly de-identifies information most of the time, it makes undermarking and overmarking errors. For example, the date represented as 2008-11-10 would miss because the use of the full-width character '-'. However, this character is useful for typesetting Latin characters in a CJK environment (a collective term for Chinese, Japanese, and Korean, which constitute the main East Asian languages). One suggestion is that the approach may be customized to work on different types of medical record. A limitation of de-identification described in this paper cannot remove all PII in a medical record perfectly in the non-defined type. Although it has a little error in non-defined EHRs, we believe that it is better than starting manually from the beginning.

Performing a manual review of de-identification results by patients is not only to reduce the error rates, but also to disclose partial PII that may also be implied from the different circumstances. For example, a physician who has diagnosed disease is directly involved in a patient's care or treatment. The patient could disclose the physician's name from his/her medical record. In this case, partial disclosure of PII by patient selection may be necessary to ensure continuation of patient care or treatment. On the other hand, patients could also seek to obtain expert information through the anonymous forums. Patients do not want to expose their all PII when they seek medical advice. The patient selection process could provide the high capability to satisfy the patients' needs for different purposes. The proposed method is customized to accommodate selecting or changing the privacy protected areas by the patients in accordance with their right to legally use their own EHRs.

The proposed method could work with support resources including smartcard, digital certificate, and TTPs under the Public Key Infrastructure (PKI) environment. Some mechanisms such as watermark and keyed-hash message authentication code (HMAC) with different devices storing the PKI-related information can also provide the integrity of EHRs. For example, the key files, which are used for digitally signing EHRs, are stored directly on a hard disk, floppy disk, smartcard, or mobile equipment, each possessing advantages and disadvantages. However, digital signature mechanisms should be incorporated into the client applications to ensure the integrity of EHRs. Therefore, these devices would be chosen to support the digital signature mechanism, considering the various features of each device such as security level, usability portability, ubiquity, cost, etc.

Interoperability is a key factor for a successful implementation of an EHR system. The major purpose of EHR standards is to facilitate improvements regarding interoperability, security, reliability, efficiency and communication. Several standards development organizations, such as International Standards Organization (ISO), European Committee for Standardization (CEN), Health Level 7 (HL7), American Society for Testing and Materials (ASTM) and DICOM, are involved in the development of EHR-related standards. Recently, the Healthcare Information Technology Standards Panel (HITSP) decided to use the combined harmonized HL7 ASTM Continuity of Care Document (CCD) [36]. If the CCD standard becomes the core standard for personal EHRs, it will permit EHRs to be both portable and interoperable with other healthcare information systems. Integrating the

Healthcare Enterprise (IHE) has introduced a standard integration profile, Exchange of Personal Health Record Content Profile (XPHR), in IHE technical frameworks. The profile specifies the operations to extract healthcare contents from hospital information systems to a personal health record (PHR). The methods proposed in this study have great potential for incorporation with the profile to protect the privacy of PHR, especially in the well-defined type such as XML-based medical records.

5. Conclusion

Portable EHRs will likely be the trend in the future, because patients want to exercise their right to access their EHRs. This means that when patients have their EHRs from hospitals, they will take over responsibility for the information obtained. In the present study, we adopt a simple and practical method for safeguarding the EHR in real world applications. The approach is able to reduce the risk of security concerns on the EHR and offers personal privacy protection. Using the automatic system we have designed, patients are allowed to control the data by adding or deleting items to be protected, which helps minimize security risks when carrying their medical data. It also mitigates concerns about leakage of personal information during the transportation of data as well as privacy when consulting with a third party expert for medical advice. The processed EHR also comes with readability. The results showed the portable exchange of EHRs by our method based on de-identification, pseudonymity, cryptographic technology and patient control is feasible and practically useful. Our method provides a solution for safeguarding EHRs, and allows patients to protect the privacy and security of their medical information.

6. Summary

In this study, the method of privacy preservation and information security protection by protecting the identifiers has been shown to be a good way to protect patients' portable EHRs. Patients are the keepers of protection of their portable EHRs. The method is a secure process (de-identification, pseudonymity, patient selection and encryption), and includes a recovery process (decryption, verification) to protect patients' EHRs when they are being held and used by patients. It can protect patients' EHRs when they are carried outside the hospital. Besides giving information security protection with smartcard by cryptography such as protecting under PKI, our method also provides for the privacy preservation of EHRs. The secured EHRs are still readable. The method is useful to avoid inappropriate, unintentional and wrongful disclosure when patients are using their EHRs, such as to obtain a second opinion through Internet by email, or in private/public website forum. It is suitable for patients who are using portable storage media and who want to have access to their EHRs with security and privacy protection.

Conflict of interest statement

None declared.

Acknowledgements

This work was supported by a grant from the National Science Council, Taiwan (NSC 95-2221-E010-003). We are grateful for the support provided by the Information Office of Changhwa Christian Hospital, Changhwa, Taiwan.

References

[1] Code of Federal Regulations (US), Title 45, Part 164, 2002.

[2] B. Sadan, Patient data confidentiality and patient rights, Int. J. Med. Inf. 62 (2001) 41–49.

[3] J.B. Fowles, A.C. Kind, C. Craft, E.A. Kind, J.L. Mandel, S. Adlis, Patients' interest in reading their medical record: relation with clinical and sociodemographic characteristics and patients' approach to health care, Arch. Intern. Med. 164 (2004) 793–800.

[4] A.K. Abdullah, Protecting your good name: identity theft and its prevention, in: Proceedings of the First Annual Conference on Information Security Curriculum Development, Kennesaw, Georgia, October 2004, pp. 102–106.

[5] C.M. Yang, H.C. Lin, P. Chang, W.S. Jian, Taiwan's perspective on electronic medical records' security and privacy protection: lessons learned from HIPAA, Comput. Meth. Prog. Bio. 82 (2006) 277–282.

[6] I.L. Horowitz, Privacy, publicity and security: the American context: privacy is not only a right but also an obligation, EMBO Report, vol. 7 (SI) 2006, pp. S40–S44.

[7] A. Hassol, J.M. Walker, D. Kidder, K. Rokita, D. Young, S. Pierdon, D. Deitz, S. Kuck, E. Ortiz, Patient experiences and attitudes about access to a patient electronic health care record and linked web messaging, J. Am. Med. Inform. Assoc. 11 (2004) 505–513.

[8] R.C. Barrows Jr., P.D. Clayton, Privacy, confidentiality, and electronic medical records, J. Am. Med. Inform. Assoc. 3 (1996) 139–148.

[9] G.I. Serour, Confidentiality, privacy and security of patients' health care information: FIGO Committee for the Ethical Aspects of Human Reproduction and Women's Health, Int. J. Gynecol. Obstet. 93 (2006) 184–186.

[10] E. Ball, D.W. Chadwick, D. Mundy, Patient privacy in electronic prescription transfer, IEEE Security & Privacy 1 (2003) 77–80.

[11] K.D. Mandl, P. Szolovits, I.S. Kohane, Public standards and patients' control: how to keep electronic medical records accessible but private, Brit. Med. J. 322 (2001) 283–287.

[12] K.D. Mandl, W.W. Simons, W.C.R. Crawford, J.M. Abbett, Indivo: a personally controlled health record for health information exchange and communication, BMC Med. Inform. Decis. Making 7 (2007) 25.

[13] S. Gritzalis, C. Lambrinoudakis, D. Lekkas, S. Deftereos, Technical guidelines for enhancing privacy and data protection in modern electronic medical environments, IEEE Trans. Inf. Technol. Biomed. 9 (2005) 413–423.

[14] C.T. Liu, P.T. Yang, Y.T. Yeh, B.L. Wang, The impacts of smart cards on hospital information systems—an investigation of the first phase of the national health insurance smart card project in Taiwan, Int. J. Med. Inf. 75 (2006) 173–181.

[15] M.I. Kim, K.B. Johnson, Personal health records: evaluation of functionality and utility, J. Am. Med. Inform. Assoc. 9 (2002) 171–180.

[16] T. Churches, A proposed architecture and method of operation for improving the protection of privacy and confidentiality in disease registers, BMC Med. Res. Methodol. 3 (2003) 1.

[17] J. Car, A. Sheikh, Email consultations in health care: 1—scope and effectiveness, Brit. Med. J. 329 (2004) 435–438.

[18] J. Car, A. Sheikh, Email consultations in health care: 2—acceptability and safe application, Brit. Med. J. 329 (2004) 439–442.

[19] D.E. Gobuty, W. Leetz, R.J. Horn, J.F. Moehrke, Allocating basic security rules for use in medical imaging information technology, Acad. Radiol. 11 (2004) 779–786.

[20] E. Meux, Encrypting personal identifiers, Health Serv. Res. 29 (1994) 247–256.

[21] P. Szolovits, I. Kohane, Against simple universal health-care identifiers, J. Am. Med. Inform. Assoc. 1 (1994) 316–319.

[22] R. Agrawal, C. Johnson, Securing electronic health records without impeding the flow of information, Int. J. Med. Inf. 76 (2007) 471–479.

[23] R. Miller, J.K. Boitnott, G.W. Moore, Web-based free-text query system for surgical pathology reports with automatic case deidentification, Arch. Pathol. Lab. Med. 125 (2001) 1011.

[24] J.J. Berman, Concept-match medical data scrubbing: how pathology text can be used in research, Arch. Pathol. Lab. Med. 127 (2003) 680–686.

[25] M. Douglass, G.D. Clifford, A. Reisner, G.B. Moody, R.G. Mark, Computer-assisted de-identification of free text in the MIMIC II database, Comput. Cardiol. 31 (2004) 341–344.

[26] D. Gupta, M. Saul, J. Gilbertson, Evaluation of a deidentification (De-Id) software engine to share pathology reports and clinical documents for research, Am. J. Clin. Pathol. 121 (2004) 176–186.

[27] B.A. Beckwith, R. Mahaadevan, U.J. Balis, F. Kuo, Development and evaluation of an open source software tool for deidentification of pathology reports, BMC Med. Inform. Decis. Making 6 (2006) 12.

[28] H. Müller, J. Heuberger, A. Geissbuhler, Logo and text removal for medical image retrieval, in: German Workshop on Medical Image Retrieval (BVM), Springer Informatik Aktuell, Heidelberg, 2005.

[29] E.M. Newton, L. Sweeney, B. Malin, Preserving privacy by de-identifying facial images, IEEE Trans. Knowl. Data Eng. 17 (2005) 232–243.

[30] K. Pommerening, M. Reng, Secondary use of the electronic health record via pseudonymisation, Medical and Care Compunetics 1, Studies in Health Technology and Informatics, vol. 103, IOS Press, Amsterdam, The Netherlands, 2004, pp. 441–446.

[31] R. Noumeir, A. Lemay, J.M. Lina, Pseudonymisation of radiology data for research purposes, Proc. SPIE 5748 (2005) 298–305.

[32] D. Appleman, Regular expressions with NET, Desaware Inc., ⟨http://www.desaware.com⟩, 2002.

[33] A. Watt, Beginning Regular Expressions (Programmer to Programmer), Wrox Press, Wiley, Indianapolis, IN, 2005.

[34] American College of Radiology, National Electrical Manufacturers Association, ACR-NEMA Digital Imaging and Communications Standard: DICOM 2008 Part 15 Security Profiles, NEMA, Rosslyn, VA, USA.

[35] N. Weiss, E-mail consultation: clinical, financial, legal, and ethical implications, Surg. Neurol. 61 (2004) 455–459.

[36] ASTM International (ASTM), ASTM E2369-05 Standard Specification for Continuity of Care Record (CCR).

Lu-Chou Huang received Ph.D. degree from the Institute of Biomedical Engineering, National Yang-Ming University in 2009. His research interests are in the area of biomedical signal and image processing, information security and medical information system.

Huei-Chung Chu is Associate Professor of Information Management at Hua-Fan University, Taipei, Taiwan. He received Ph.D. in Computer Science from Illinois Institute of Technology, USA (1990). His current research interests are in the field of information security management, IT service management, IT governance and computer forensics.

Chung-Yueh Lien received B.S. degree in Electronic Engineering from Fu-Jen Catholic University in 2002, and M.S. degree in Biomedical Engineering, National Yang-Ming University in 2004. He is particularly interested in information security, integration of EHR, DICOM, and IHE.

Chia-Hung Hsiao is Assistant Professor of Medical Informatics at Tzu Chi University (August 2007—present). He received Ph.D. degree from the Institute of Biomedical Engineering, National Yang-Ming University in 1999. He was an Information Administrator in Computer and Communication Center, National Yang-Ming University (April 2001–July 2007). His research interests are in information security, PACS, IHE, and e-learning for medical education.

Tsair Kao received Ph.D. degree in Electrical Engineering from the University of Michigan in 1984. He was a professor at National Yang-Ming University from 1989 to 2008. Since 2009 he has been a Professor of Biomedical Engineering at Hungkuang University. His research interests include automated analysis of bioelectrical signals, digital signal processing, and computer applications in medicine.