코드화된 건강기록의 항목별 민감도에 따른 프라이버시 보호 방법

도형호*, 이성기**
*경북대학교 전자전기컴퓨터학부
*경북대학교 IT대학 컴퓨터학부
e-mail:sklee@knu.ac.kr

A Privacy Protection Method for Coded Health Record focused on the Sensitivity for Each Element

Hyoungho Do*, Sungkee Lee*
*School of Electrical Engineering and Computer Science, Kyungpook National University
*School of Computer Science and Engineering, Kyungpook National University

요 약

개인건강기록에서 프라이버시를 보호하기 위한 방안으로 환자의 식별정보를 제거하는 익명화와 식별정보를 가상의 식별자로 변경하였다가 권한을 가진 사용자가 열람할 수 있게 복원하는 가명화, 그리고 건강기록을 암호화하여 정보를 보호하는 방법들이 연구되어 왔다. 본 논문에서는 용어표현을 위해 국제표준코드를 사용하는 건강기록에서 항목별 정보의 민감도에 따라서 암호화 수준을 달리하여 정보전체를 암호화하는 것 보다 효율적이고 강력한 보안수준을 유지할 수 있는 방법을 제시한다.

1. 서론

오늘날 전자화된 의료정보의 교류는 의료인의 의사결정과 의료비용을 감소시키는 등의 긍정적인 효과들을 내고 있다. 미국에서는 S-EHR(Sharable Electronic Health Record) 시스템의 도입을 통하여 연간 1750억 달러 이상의 약물부작용을 방지하는 경제적 효과와 약물부작용으로 인한 사고를 15만건 이상 감소시키는 것으로 추정하고 있다[1]. 의료정보의 교류는 언제 어디서나 쉽게 개인건강기록에 접근할 수 있는 이점이 있는 반면에 개인건강기록을 누구나 쉽게 열람할 수 있어 프라이버시를 침해할 수 있다는 단점이 존재하며, 최근 사회적으로 이러한 개인건강정보의 오남용에 대한 규제와 목소리가 높아지고 있다[2].

프라이버시 보호를 위한 방법으로 개인식별정보를 제거하여 누구의 건강기록인지 식별하지 못하게 하는 익명화(anonymization), 개인식별정보를 가상의 이름으로 변경시키고 식별정보를 따로 보관하였다가 필요시 복원하는 가명화(pseudonymization), 그리고 환자의 건강기록을 본인 혹은 주치의와 같이 허가된 사용자만 복호화(decryption)하여 열람할 수 있도록 암호화(encryption)하는 방법을 중심으로 연구되어 왔다.

하지만 이러한 방법들은 각각의 단점들을 가지고 있다. 익명화는 환자의 식별자를 제거하는 것을 기본으로 하기 때문에 의료행위를 목적으로 특정환자의 건강정보임을 식별할 수 없어서, 연구자들이나 보험사 등의 조사 목적으로 주로 사용되고 있다. 특정 환자의 진료의 목적으로 사용되는 건강기록의 경우는 개인식별정보를 가명화(pseudonymization)하고 실제 환자정보를 따로 보관함으로써 열람의 권한을 가지는 사람에게만 개인식별정보를 포함한 정보를 제공할 수 있게 하는데, 가명화된 건강정보는 프라이버시 침해의 위험은 줄어드나, 개인식별정보 저장소가 공격의 대상이 될 수 있는 문제점에 노출되는 문제점이 발생한다. 건강기록을 암호화를 시키는 방법은 가명화보다는 안전하게 정보를 보호할 수 있는 방법이 될 수 있으나 비밀번호가 노출되거나 의료기관 내부자를 통한 환자 프라이버시가 불필요하게 노출될 수가 있다. 또한 건강기록 중 수백 메가바이트 이상의 데이터도 존재하며, 이 경우 복호화시 많은 시간이 소요되어 적절한 시간에 의료정보를 열람할 수 없는 문제점이 발생할 수 있다[3]. 본 논문에서는 HL7 CDA[4]와 같이 국제표준코드를 사용하여 구성되는 용어를 표현하는 건강기록에서 항목별 민감도에 따른 암호화 수준을 차별화하는 방법을 제안한다. 본 논문에서 제안한 방법은 다음과 같이 공헌한다. 개인건강기록에서 민감한 정보만을 암호화함으로써 전체정보를 암호화하는 것보다 시스템 자원을 절약하고 암·복호화를 위한 시간을 단축한다. 또한 각 항목의 민감한 정도에 따라서 암호화의 수준을 차별화함으로써 효율적이고 건강정보문서 전체를 암호화하는 것 보다 민감한 정보에 한해서 더 높은 보안수준을 유지할 수 있도록 한다.

2. 관련연구

2.1 의료정보의 프라이버시 보호 방법

익명화는 조사 혹은 연구를 위한 마이크로 데이터에서 프라이버시를 보호하기 위하여 제안된 방법이다. 익명화는 원본 데이터로부터 그룹화를 기반으로 프라이버시를 확보한다. 데이터 내에 특정자료와 식별이 불가능한 데이터 목록이 적어도 k-1개 존재하도록 데이터를 변형하여 프라이버시를 확보하는 k-익명성(k-anonymity)을 기본으로 하며[5], 목록 내에서 민감한 속성 값들이 모두 같거나 다양하게 구분되지 않는 경우에 사용자 식별정보를 추정할 수 있다는 문제점에 대한 해결책으로 l-다양성(l-diversity)[6], 그리고 k-익명성과 l-다양성을 만족하더라도 민감한 속성의 유사도가 높으면 정보를 추측할 수 잇다는 문제점과 그에 대한 해결책을 제시한 t-밀접성(t-clossness)[7]이 차례로 제안되었다. 익명화는 기본적으로 식별 정보를 제거하여 다시 복원하지 않기 때문에 특정 환자의 건강정보를 열람하는 목적이 아닌 조사 혹은 연구 목적으로 사용할 때는 높은 수준의 프라이버시 보호를 할 수 있으나, 특정 환자의 건강정보를 열람할 수 없다는 단점이 존재한다.

가명화는 건강기록내의 식별자를 가상의 식별자로 변경하고 원본의 식별자는 별도로 보관하는 방법으로 불법적인 공격에 대한 문제점과 해결방법에 관한 것과 건강기록과 식별자사이의 페어(pair) 정보를 보호하는 방안에 관해서 주로 연구되어 왔다[8][9].

암호화는 건강기록을 다양한 암호화 알고리즘을 통하여 암호화시키고 그 키(key)를 가지고 있는 사람에 한해서 건강기록을 복호화하여 열람할 수 있도록 하는 방법이며, 의료영역에서는 주로 건강정보의 암호화를 효율적으로 하는 방법과 건강정보를 휴대하기 위하여 암호화를 적용시키는 방법 등의 다양한 연구가 이루어 졌으나 의료정보의 분석을 통한 세부적인 암호화에 관한 접근은 미비하였다[10][11].

그 외 IHE XDS(Integrating the Healthcare Enterprise) XDS(Cross Enterprise Document Sharing)환경에서 HL7 CDA의 헤더정보와 바디정보를 서로 다른 저장소(repository)에 저장함으로써 불법적인 공격에도 누구의 건강기록인지 알 수 없도록 하는 의료정보표준에 특성화된 연구가 진행되었다[12].

2.3 의료용어 표현을 위한 국제 표준 코드

의료 용어표현을 위한 대표적인 국제표준코드는 다음과 같다.

- LOINC(Logical Observation Identifiers Names and Codes): 검사정보의 공유와 표준화를 위한 국제표준으로 로 검사명과 검체, 검사방법. 등이 포함된 6개 항목의 조합으로 구성된 코드
- ICD-9-CM(The International Classification of Diseases, Ninth Revision Clinical Modification): 질병 및 관련 건강 문제의 국제 통계 분류 9차 개정판으로 세계 보건 기구에서 질병과 증상 등을 분류해놓은 것
- ICD-10(The International Classification of Diseases, tenth) : 질병 및 관련 건강 문제의 국제 통계 분류(ICD) 10차 개정판으로, 세계 보건 기구에서 질병과 증상 등을 분류해놓은 것
- SNOMED(systematized Nomenclature of Medicine) : 의료인이 질병, 임상소견, 처치 등을 표현하기 위한 포괄절 임상용어이며 질병의 이름 혹은 환자의 상태를 묘사할 수 있도록 제공
- UMLS(Unified Medical Language System): 미국 국립의학도서관의 연구개발팀이 다양한 정보원으로부터 얻은 정보를 통합하고 검색할 수 있는 시스템을 구축하기 위해 개발한 용어체계
- 그 외

  NANDA, NINDA, ICNP: 간호 용어

  EAN/UCCC-13, ATC: 의약품 용어

  UNSPSC: 의료재료 용어

3. 용어코드 종류에 따른 민감도 테이블

개인의 건강기록에서 가장 민감도가 높은 정보는 병명에 관한 정보이다. 또한 같은 병명이라도 약물중독, 성병, 정신질환 등과 관련된 질병은 더욱 민감한 정보이다.

질병을 분류를 나타는 코드로는 ICD-9CM, ICD-10이 있고 SNOMED 역시 질병을 나타내는 코드가 포함되어 있다. 검사에 관한 정보는 주로 LOINC를 사용하며 질병의 종류를 나타내는 정보보다는 민감도가 떨어지나 HIV/AIDs 등의 민감한 질병을 검사하는 코드는 다른 검사정보에 비해 더 민감 할 수 있다. UMLS의 경우는 의학도서관을 위한 용어로서, 코드값에 따라서 그 민감도가 달라질 수 있다. 그 외 의약품용어나 간호용어, 의료재료 용어 등은 위에서 언급한 용어들에 비해 민감도가 떨어짐을 알 수 있다. 본 논문에서는 각 용어의 종류 및 해당코드에 따라서 민감도에 관한 가중치 테이블을 사용하여 암호화의 수준을 결정할 수 있도록 하는 시스템을 제안하며, 가중치 테이블의 예는 표 1과 같다.

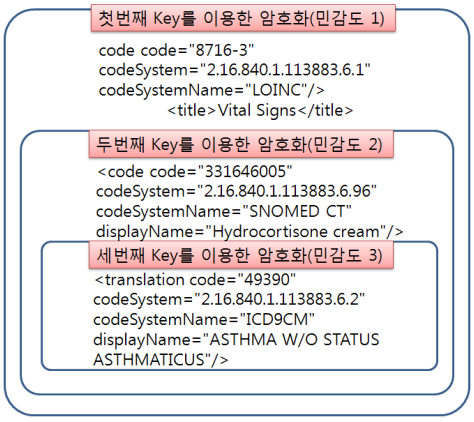
<표 1> 용어코드에 따른 민감도 가중치 테이블의 예

Code System OID	Code System	Code	Sensitivity
2.16.840.1.113883.6.1	LOINC	8716-3	1
2.16.840.1.113883.6.96	SNOMED-CT	331646005	2
2.16.840.1.113883.6.2	ICD9CM	*	3
2.16.840.1.113883.6.2	ICD9CM	042	4

각 코드별 민감도를 설정하여 암호화 수준을 결정하기 위하여 표 1 같은 정보를 미리 설정하여 사용한다. 병명을 나타내는 ICD-9CM코드에 모든 코드- 와일드카드(*)를 사용-는 민감도가 3으로 되어 있으나, ADIS를 표현하는 ICD-9CM코드의 경우는 민감도가 4로 설정하여 민감도가 다른 질병에 비해 높게 설정되어 있는 것을 볼 수 있다. 이는 해당 의료진이나 보안담당자가 상황을 고려하여 편집 하여 암호화 시스템에 가중치로 사용될 수 있다.

4. 민감도 가중치에 따른 암호화 시스템

본 논문에서는 그림 1과 같이 정의한 민감도 가중치 테이블 값에 따라서 암호화 횟수를 달리하여 정보의 보호수준을 달리하는 방법을 사용한다. 그림 1과 같이 "vital sign"에 관한 정보는 민감도가 1이기 때문에 1회의 암호화만 수행하고 "천식"이라는 병명을 가지고 있는 ICD-9CM코드의 경우 총 3회의 암호화를 거치게 된다.



<그림 1> 민감도 가중치에 따른 암호화

3회의 암호화를 거치는 동안 각 단계별 암호화는 각각의 키를 사용하며, 해당 환자가 천식이라는 병을 가지고 있다는 정보를 얻기 위해서는 3개의 패스워드를 각각 사용하여 3회의 복호화를 수행하여야 하며, 이는 민감한 정보일수록 강력한 보안수준을 가질 수 있음을 보여준다.

5. 결론

본 논문에서는 용어표현을 위해 국제표준코드를 사용하는 건강기록에서 항목별 정보의 민감도에 따라서 가중치 수준을 결정하여 항목별로 암호화하는 방법을 제시하였고, 정보전체를 암호화 하는 것 보다 효율적이고 민감한 속성에 대해서는 더 강력한 보안수준을 유지할 수 있는 방법을 제안하였다. 하지만 논문에서 제안한 내용은 코드화된 정보에 대해서만 언급하고 일상적인 서술(narrative statements) 에 대한 부분은 언급하지 않았다. 또한 대표적으로 많이 사용하는 코드값에 대한 예를 들었을 뿐 코드별로 구체적인 분석이 이루어지지는 않았다. 향후 연구과제로는 이러한 일상적인 서술과 다양한 용어코드에 대한 분석을 통하여 실제에 활용할 수 있는 시스템을 구현하는 것이다. 그리고 AIDS와 같은 병명을 가지는 최고민감도를 가지는 정보의 경우는 병명을 구체적으로 알 수는 없지만 매우 민감한 병을 가지고 있다는 사실을 추측할 수 있게 한다. 이를 해결하는 것도 향후 연구과제이다.

참고문헌

[1] Frank R. Ernst and Amy J. Grizzle, "Drug-Related Morbidity and Mortality: Updating the Cost-of-Illness Model", Journal of the American Pharmaceutical Association, Vol. 41, No 2, pp.192-199, Mar. 2001.

[2] 김운석, "개인정보보호 2.0시대의 개인정보보호법 개관", 충남대학교 법학연구, Vol. 22, No.2, pp.9-41, Dec, 2011

[3] J. Montagnat et al, "Medical images simulation, storage, and processing on the European DataGrid Testbed", Journal of Grid Computing , Vol 2, No.4, pp. 387 - 400, Jul, 2004

[4] Robert H. Dolin et al, "HL7 Clinical Document Architecture, Release 2", J Am Med Inform Assoc Vol 13, No.1, pp.30-39, 2006

[5] L. Sweeney "*k*-anonymity : A model for protection privacy," International Journal of Uncertainty, Fuzziness and Knowledge-based systems, Vol. 10, No. 5, pp. 557-570, October 2002.

[6] A. Machanavajjhala et al, "*l*-diversity: Privacy beyond *k*-anonymity", Proceedings of the International Conference on Data Engineering (ICDE2006), pp. 24-35, April 2006.

[7] T. Li et al "*t*-closeness: Privacy beyond *k*-anonymity and *l*-diversity," Proceedings of the International Conference on Data Engineering(ICDE2006), pp.106-115, April 2006.

[8] Bradley A. Malin, "An Evaluation of the Current State of Genomic Data Privacy Protection Technology and a Roadmap for the Future", J Am Med Inform Assoc, Vol 12, No 1, pp.28-34, Jul 2005

[9] Thomas Neubaue et al, "A methodology for the pseudonymization of medical data", Int J Med Inform. Vol 80, No 3, 190 - 204, Nov 2011.

[10] Adam Wright et al, "Encryption Characteristics of Two USB-based Personal Health Record Devices" J Am Med Inform Assoc, Vol 14, No 4, Jul 2007.

[11] Huang KH et al, "Application of portable CDA for secure clinical-document exchange.", Journal of Medical Systems, Vol 34, No 4, pp.531-539, Feb 2009.

[12] 김일광, "익명화 방법을 적용한 임상진료문서 등록 기법 연구", 정보과학회논문지 Vol 34, No 10, pp.918-928 Oct 2007.