An Efficient Approach for Preserving the Medical Data using Homo Morphic Encryption

P. Saranya[1], C. Usha Nandhini[2]

[1]M.Phil Research Scholar
Department of Computer Science,
Vellalar College for Women (Autonomous), Erode-12, India
saran3.saranya@gmail.com

[2]Assistant Professor
Department of Computer Application
Vellalar College for Women (Autonomous), Erode-12, India
cushanandhini@yahoo.co.in

Abstract: ***An electronic medical record (EMR) is a computerized medical record created in an organization that delivers care, such as hospital or lab. Emerging policies encourage investigators to disseminate such data in a de-identified form for reuse and collaboration, but organizations are hesitant to do so because they fear such actions will jeopardize patient privacy. The two techniques suppression-based and generalization-based k-anonymous databases are used to preserve patient's privacy but reidentification is possible to break that privacy. This paper proposes a well-known cryptographic assumption, a homomorphic encryption that offers patients privacy and the details in a de-identified form. We demonstrate that the proposed approach can generate anonymized data that permit effective biomedical analysis using several patient cohorts derived from the EMR System.***

Keywords: Electronic medical records, Privacy, Homo morphic Encryption, Anonymous.

1. Introduction

Advances in health information technology have facilitated the collection of detailed, patient-level clinical data to enable efficiency, effectiveness, and safety in health-care operations. Such data are often stored in electronic medical record (EMR) systems and are increasingly re-purposed to support clinical research. The problem is to formally anonymize longitudinal patient records. Methods to mitigate re-identification via demographic and clinical features are not applicable to the longitudinal scenario. These methods assume the clinical profile is devoid of temporal or replicated diagnosis information. Consequently, these methods produce data that are unlikely to permit meaningful longitudinal investigations. So the methods for preventing re-identification in relational data (e.g., demographics) are reviewed, where records have a fixed number of attributes and one value per attribute.

2. Existing Methodology

If the information for each person contained in the release cannot be distinguished from at least k-1 individuals whose information also appears in the release. For example, if they try to identify a man from a release, but the only information they have is his birth date and gender. There are k people meet the requirement. This is called k-Anonymity.

2.1 Existing Techniques

The K- anonymity techniques are Generalization and Suppression. In generalization-based Anonymization method, original values are replaced by more general values, according to a priori established Value generalization hierarchies (VGHs). In suppression-based Anonymization method, mask with the special value, the value deployed by (database owner) for the Anonymization. Our approach can be extended to prevent this attack by controlling generalization and suppression to ensure that an additional principle is satisfied, such as ℓ-diversity which dictates how sensitive information is grouped. But re-identification is not controlled by using these techniques, and it does not limit the amount of information loss incurred by generalization and suppression.

Table 1: Generalized & Suppressed Dataset

S.No	Blood Test Name	Blood Group	Age	Disease
Original Dataset				
1	Amylase Blood Test	A1B+	34	Cancer
2	HIV Blood Test	B+	53	AIDS
Generalized Dataset				
1	Blood Test	A1B+	30-40	Cancer
2	Blood Test	B+	40-50	AIDS
Suppressed Dataset				
1	*	A1B+	*	*
2	*	B+	39	*

3. Proposed Methodology

Homomorphic encryption is used to preserve the medical data in de-identified form. It is a form of encryption which allows specific types of computations to be carried out on cipher text and obtain an encrypted result which is the cipher text of the result of operations performed on the plaintext. Homomorphic encryption is expected to play an important part in cloud computing, allowing companies to store encrypted data in a public cloud and take advantage of the cloud provider's analytic services. Data loss and data redundancy also reduced. The process of Homomorphic Encryption contains the following steps:

1. Charlie codes his tuple δi into c(<t1',... ts'>), is denoted as c(δi). Then, he encrypts c (δi) with his private key and sends EA c (δi) to Dora.

2. Dora separately codes all attribute value in V to get the coded tuple values < C (t1), C (tU) >, encrypts each coding and EA (c(δi)) with her key B and sends (i) < EB (C (t1),...C (tU)) >; and (ii) EB (EA (c (δi))) to Charlie.

3. Since E is a commutative encryption scheme,

EB (EA (C (δi))) = EA (EB (C (δi)))
Charlie decrypts,
EA (EB (C(δi))) = EB (EA (C (δi))) to obtain EB (EA (C (δi))).

4. Since the encrypted values send by Dora are attributes in R, Charlie knows the encrypted values send by Dora, the one corresponding to the suppressed and non suppressed attributes. Thus, Charlie computes

EB (C (t1) X …………X EB (C (tS)

Where v1; ...; vs are the values of nonsuppressed attributes contained in tuple t. As already mentioned, E is a product-homomorphic encryption scheme based also on the definition of function C (.), this implies that Expression (A) is equal to

EB (C (< t1 …… tS >))

5. Charlie checks whether

EB (C (< t1 …… tS >)) =
EB (C (< t'1 …… t'S >))

If the condition is true, the attribute value of V will be inserted to the database table R, else it breaks the k-anonymity rules. Secure delivery of medical data to and from the cloud is however a serious issue that needs to be addressed. The security issues affecting cloud computing and propose the use of homomorphic encryption as a panacea for dealing with these serious security concerns vis-à-vis the access to cloud medical data.

Table 3: Homomorphic Encryption

S.No	Dataset Name	Size of Dataset (In Mb)	Key Generation (In Sec)	Encryption (In Ms)	Decryption (In Ms)
1	Heart dataset	16	0.16	4	4
2	Lung dataset	20.5	1.00	7	10
3	Blood dataset	24	1.25	13.5	23

4. Result and Discussion

In Proposed work, the data loss is minimally controlled by using the Homomorphic encryption scheme. The ability to perform simple deterministic computations on encrypted data make homomorphic cryptosystems ideal for creating privacy preserving protocols. In general, the protocols presented are meant to be general building blocks for further applications. For example, utilizing homomorphic cryptography to perform simple set operations provides tools that can be used to construct even more complicated protocols, such as complicated database queries.

Table 2: Comparison of Generalization and Suppression Techniques with Homomorphic Encryption Scheme

S.NO	TEST NAME	LOSS METRIC	
		GENERALIZATION AND SUPPRESSION (IN %)	HOMOMORPHIC ENCRYPTION (IN %)
1	Amylase Blood Test	0.95	0.03
2	TSH Blood Test	0.55	NIL
3	HIV Blood Test	0.85	NIL
4	ALT(SGPT)Blood Test	0.75	0.01
5	Blood Chemistry Test	0.65	NIL

5. Conclusion and Future Work

In this paper, procedures are carried out for privately checking whether a k- anonymous database retains its anonymity once a new tuple is being inserted to it. Since the proposed protocols ensure the updated database remains k-anonymous, the results returned from a user's (or a medical researcher's) query are also k-Anonymous. Thus, the patient or the data provider's privacy cannot be violated from any query. As long as the database is updated properly using the proposed protocols, the user queries under the application domain are always privacy-preserving. In future, the definition of a mechanism for actually performing the update, once k-anonymity has been verified and the integration with a privacy-preserving query system. Then implementing a real-world anonymous database system. Improving the efficiency of protocols, in terms of number of messages exchanged and in terms of their sizes, as well.

References

[1] B. Pinkas, "Cryptographic techniques for privacy - preserving data mining", ACM Special Interest Group on Knowledge Discovery and Data Mining Explorations, vol. 4, no. 2, pp. 12–19, 2002.

[2] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, "Privacy - preserving data publishing: A survey of recent developments," ACM Computing Surveys, vol. 42, no. 4, pp. 1–53,2010.

[3] L. Sweeney, "k-anonymity: A model for protecting privacy," International Journal of Uncertainty, Fuzziness and Knowledge – Based Systems, vol.10, no.5, pp. 557–570, 2002.

[4] Aldermen A. Atayero, Oluwaseyi Feyisetan, "Security Issues in Cloud Computing: The Potentials of Homomorphic Encryption" Journal of Emerging Trends in Computing and Information Sciences, ISSN 2079-8407.

[5] B. C. Chen, D. Kifer, K. Le Fevre, and A. Machanavajjhala, " Privacy-preserving data publishing," Foundations and Trends in Databases, vol. 2, no. 1-2, pp. 1– 167, 2009.

[6] Alberto Trombetta, Wei Jiang," Privacy-Preserving Updates to Anonymous and Confidential Databases", IEEE Transactions on Dependable and Secure Computing, Vol. 8, No. 4, July/August 2011

Author Profile

P. Saranya received the B. Sc & M. Sc degrees in Computer Science in 2009 &2011, respectively from Periyar University, Salem. She is currently doing M. Phil degree in Computer Science from Bharathiar University, Coimbatore, India.