Review Article Allied Sciences-Health Care

International Journal of Pharma and Bio Sciences

ISSN
0975-6299

A REVIEW OF RECENT ADVANCES IN PRIVACY PRESERVATION IN HEALTH CARE DATA PUBLISHING

BHANUMATHI SELVARAJ*[1] AND SAKTHIVEL PERIYASAMY[2]

[1]Department of Computer Science and Engineering, Sathyabama University, India
[2]Department of Electronics and Communication Engineering, Anna University, India

ABSTRACT

Nowadays, Electronic Health Records (EHRs) sharing is extremely helpful for medical data analysis while preservation of patients' privacy plays a major role. Several methods have been proposed to preserve privacy during data publishing without more utility loss. This paper reviews the current literature on privacy preserving health care data publishing. It mainly focuses on the recently (2009-2016) proposed methods based on anonymization and encryption, and their limitations to the privacy attacks. We also conclude this review with some future research direction in this field.

KEYWORDS: **Privacy, electronic health records, anonymization, encryption**

✉ **BHANUMATHI SELVARAJ**
Department of Computer Science and Engineering, Sathyabama University, India

INTRODUCTION

Privacy Preserving Data Publishing (PPDP) offers the methodology to publish data that are useful for many research purposes while preserving data privacy.[1] In the medical field, lots of health data are available that are used by several researchers for data analysis. Traditionally, health data have been recorded in the white paper. These paper health records have underrated compared to computerized health records due to the following reasons: unreadable, inadequate and unwell organized, hard to guarantee the quality of care. By the advancement of computer technology, the huge possibilities are offered for Electronic Medical Records (EMR) documentation and their use in the field of medication management.[2] The widespread usage of Electronic Health Records (EHR)[3,4]/EMR[5] system accumulates more and more health care data of patient's, for instance, name, age, gender, diagnosis codes, test results, medication, radiology images and total charge, from various sources. Sharing of these data is extremely beneficial to medical studies, clinical trials, scientific and commercial research. However, health care data usually have a huge quantity of person-specific and sensitive data of the patients, which must be protected from various privacy attacks during data sharing.[6] Also, it must be useful for successive data analysis. To extract useful information from data analysis, the huge amount of health data must be integrated from various sources. For example, Palaniappan & Huey[7] developed a tool for integrating health data from different healthcare providers in Malaysia. There is the possibility of data leakage during data integration and data outsourcing.[8] Dubovitskaya et al.[9] proposed eHealth architecture to efficiently integrate health data with no disclosure of sensitive data. In recent years, the information privacy and security issues in the health care sector have gained great attention. To balance patients privacy and data sharing, some policy mechanisms are discussed.[10] The re-identification attacks on health information is systematically analysed.[11,12] Recently, Kho et al.[13] implemented a secure tool to minimize re-identification risk in EHR data over multiple sites in US. Several studies have been discussed privacy and security concerns in health care data publishing.[14-23] However, this review focuses on novel anonymization and encryption methods proposed for secure publication of health care data in the recent (2009-2016) research works. This paper is organized as follows. We define the preliminary concept in Section 2. We present various anonymization methods that help to protect health care data in Section 3. In Section 4, we discuss how encryption methods used to solve privacy issues. Section 5 concludes this review with future research direction.

PRELIMINARY CONCEPTS

Privacy Operations

Anonymization

Anonymization is the kind of sanitization process which protects the private information of individuals and ensures privacy principles like *k*-anonymity[24-26], *l*-diversity[27], *ε*-differential privacy[28,29] using generalization and suppression, microaggregation, etc.[30] In the anonymization, the attributes of a given table are characterized as three types: Unique identifier attributes are used to identify individuals. Quasi-identifier attributes are combined with the external source to re-identify individual records. Sensitive attributes cannot be allowed an adversary to distinctively associate their values with a unique identifier.[31]

Encryption

Encryption is the tool to ensure the privacy of sensitive medical data by converting the data from one form into another. A cryptography algorithm is used with linear equation and delta encoding to protect sensitive data from unauthorized users.[32] Most important encryption algorithm for privacy preservation is homomorphic encryption. Homomorphic encryption scheme encrypts data into ciphertext that can be analyzed and used for computation without decrypting the data and, without even needing the decryption key. Also, it allows complex mathematical operations to be carried out on ciphertext without compromising the encryption. Rivest et al.[33] have initially proposed the idea of using homomorphic encryption. Gentry[34] reported the first encryption scheme, i.e. fully homomorphic encryption (FHE), which performs multiplication and addition operations. Then, it has been used to protect various data before mining.[35-38]

Generalization and suppression

Generalization refers to recoding or replacing a value with a less specific, but semantically consistent value.[39] Different types of generalization methods have been used to protect privacy of data.[40-43] Suppression involves not releasing a value at all but replaces the individual quasi-identifier attributes with a *.[39,44-46] Generalization and suppression are combined used to achieve *k*-anonymity.[39,47,48]

Microaggregation

Microaggregation is performed by two phases, data partitioning and partition aggregation. First, small clusters are constructed for the original dataset in which all cluster must have k to 2k elements. Second, all original data are replaced by the centroid of the corresponding cluster. If k is larger, the disclosure risk is lesser but information loss is larger.[49-51] Microaggregation is used for protecting various data such as categorical[52], numerical[53], and query log data.[54]

Privacy Threats

k-anonymity

The *k*-anonymity needs individual tuple should not be uniquely identifiable from a group of k on the quasi-identifier attributes. An equivalence class is referred as the set of all tuples in the table that has indistinguishable values for the quasi-identifier attributes. The table is called k-anonymous when all tuples are in an equivalence class of size at least k.[31] A conditional form of *k-anonymity* is k^m-anonymity that guarantees that partial knowledge about a tuple *r* should not be discriminate it from *k-1* tuples in the published data.[55]

l-diversity

The *l*-diversity is an extension of k-anonymity. To avoid the homogeneous sensitive data disclosed to the group, it needs every equivalence class must have at least *l* well-represented distinct values for a sensitive attribute.[31] Different variants of *l*-diversity are employed in the existing anonymization algorithms.[56,57]

LKC-privacy

The *LKC*-privacy guarantees that each combination of values in quasi-identifiers with maximum length *L* in the data table *T*. It is shared by at least *K* records, and the confidence of inferring any sensitive values in *S* is not greater than *C*, where *L*, *K*, *C* are thresholds and *S* is a set of sensitive values stated by the data owner. *LKC*-privacy bounds the probability of a successful identity linkage to be $\leq 1/K$ and the probability of a successful attribute linkage to be $\leq C$, given that the prior knowledge of adversaries does not exceed *L*.[58]

ε-differential privacy

The ε-differential privacy is defined as a randomized function *K* gives ε-differential privacy if for all data sets *D* and *D'* differs on at most one row, and all $S \subseteq$ Range(*K*).[28] It ensures the strongest privacy and makes no assumptions about background knowledge of an adversary's compared to existing privacy mode.[59] In the varying data publishing, differential privacy model has been used to prevent membership disclosure in various data publishing.[60-62]

$$Pr[F(DB1) \in SDB] \leq exp(\varepsilon) \times Pr[F(DB2 \in SDB)] \qquad (1)$$

ANONYMIZATION METHODS

Many operations such as perturbation, bucketization and slicing have been used to preserve privacy in anonymization methods. Recently, generalization and suppression, and microaggregation are more often used for protecting health care data.

Using Generalization and Suppression

The privacy anxieties of the blood transfusion information-sharing system are studied and proposed a new *LKC*-privacy model with anonymization algorithm, as listed in Table 1 (Mohammed et al., 2009).[58] The privacy-aware information sharing (PAIS) algorithm is presented to achieve *LKC*-privacy. The efficiency and scalability of this algorithm are assessed by varying the thresholds of maximum knowledge of adversary's *L*, lowest anonymity *K*, and maximum confidence *C*. The experiment on blood dataset hold by health institute and adult dataset[63] revealed that this model retains the necessary information for data analysis and scalable for anonymizing high-dimensional datasets. Gardner & Xiong[31] presented a framework for de-identifying heterogeneous health data including unstructured and structured data. The authors have empirically analyzed the Bayesian classifier, sampling based methods and conditional random fields based methods to extract sensitive data from unstructured data. Then, the data suppression and generalization have been performed to anonymize the data with various options including full/partial de-identification and statistical anonymization based on *k*-anonymization. Loukides et al.[64] illustrated the risk of a de-identified version of Vanderbilt's patient-specific data re-identification. This study estimates the level of protection and data utility of internal classification of disease (ICD) version 9 codes by applying suppression and generalization methods. They showed that existing privacy protection methods are not providing sufficient protection and valuable result in the case of complex clinical data sharing. Tamersoy et al.[5] introduced a method for securely sharing longitudinal patient-specific data. To preserve data utility, they employed alignment using generalization and suppression (A-GS) algorithm that uses dynamic programming for constructing anonymized trajectory. Then, they also used an MDAV algorithm that has the clustering component to produce anonymized data. As a result, the proposed method shares anonymized data for biomedical analysis with less information loss. The differentially private anonymization algorithm is proposed based on the generalization (DiffGen) for publishing health data which offers ε-differential privacy guarantee.[65] This algorithm first generalized the original data and then added noise to achieve ε-differential privacy. The experiment on MIMIC[66] and adult[67] datasets showed that this algorithm provides better flexibility to perform the classification analysis, and it leads trade-off among privacy protection and data utility due to the information loss by generalization method. An analytical cost model is offered which guides health information custodian's (HIC's) to make better decisions on the optimal value of releasing patient-specific health data.[68] This model is used to identify an optimal trade-off among privacy and data utility regarding monetary value. The extensive experiments on adult dataset showed that their model is helpful for HICs to attain the optimal value by selecting various privacy models, for example, ε-differential privacy, *LKC*-privacy and *k*-anonymity, under different privacy parameters and different anonymization algorithms [DiffGen and top-down specialization (TDS)]. In addition, this model might be suitable for anonymization methods (non-perturbative/perturbative) and thus this method is utilized for other types of data publishing scenarios. Loukides et al.[55] presented an approach that develops an effective disassociation-based algorithm to anonymize diagnosis codes for preventing re-identification. To preserve privacy and data utility better than other methods, this algorithm performs three operations such as vertical partitioning (VERPART) and horizontal partitioning (HORPART) and refining. The experiment on INFORMS dataset (https://sites.google.com/informsdataminingcontest/) showed that this approach is better than clustering-based anonymizer (CBA)[11] in terms of data utility, efficiency and scalability. Besides, the published data using this approach also allow different clinical case studies and medical analysis tasks.

Table 1
Summary of privacy preserving health care data publishing using anonymization methods

S. No	References	Privacy models	Privacy threats	Operations	Methods/ Algorithms	Datasets used	Comparison to other methods
1	Mohammed et al. (2009)[58]	*LKC*-privacy	Identity and attribute linkage	Generalization	PAIS	Blood and adult dataset	Not applicable
2	Gardner and Xiong (2009)[30]	*k*-anonymity	Identity disclosure	Suppression and generalization	Conceptual framework	Real cancer dataset	Naive Bayes and CRF approach
3	Tamersoy et al. (2012)[5]	*k*-anonymity	Identity disclosure	Generalization and suppression	A-GS and MDAV	BioVU and VUMC	Not applicable
4	Mohammed et al. (2013)[65]	*ε*-differential privacy	Membership disclosure	Generalization	DiffGen	MIMIC and adult	DiffP-C4.5 and TDS approach
5	Khokhar et al. (2014)[68]	*k*-anonymity, *LKC*-privacy, and *ε*-differential privacy	Identity and attribute disclosure	Generalization	DiffGen and TDS	Adult dataset	Not applicable
6	Loukides et al. (2014)[55]	k^m-anonymity	Identity disclosure	Generalization	Disassociation, VERPART and HORPART algorithms	INFORMS	CBA
7	Heatherly et al. (2014)[69]	*k*-anonymity	Identity disclosure	Generalization	phenome-wide association study	VUMC, BioVU and Demo dataset	Not applicable
8	Kim et al. (2014)[70]	*l*-diversity	Attribute disclosure	Generalization	DF anonymization	Adult and NPS	Accumulation-based methods
9	Martínez et al. (2013)[76]	*k*-anonymity	Identity disclosure	Microaggregation, recoding, resampling	SDC	OSHPD clinical dataset	Not applicable
10	Gal et al. (2014)[51]	*k*-anonymity	Identity disclosure	Microaggregation	Improved-Condensation	Real colon and lung cancer	Condensation and TFRP
11	Wang et al. (2015)[70]	*ε*-differential privacy	Attribute disclosure	Data partition and Generalization	Optimal and Greedy sanitization	Checkuplist1 and Checkuplist2	MulAnony and MulDiff

Heatherly et al.[69] employed *k*-anonymization method for simulations of the data protection process of streams structured data in the EMR. They analyzed how anonymizing various size of EMR data affected the correlation of genome-phenome association strengths and indicated that the result of large-scale data anonymization retains better utility as that of small-scale data. Kim et al.[70] suggested delay-free (DF) anonymization method to preserve the secrecy of electronic health data. This method minimizes the delay incurred during the process of data streams. The efficiency of this method is evaluated using the national patient's sample (NPS) and adult datasets, and revealed that their method significantly reduces counterfeit values and increases the utility of anonymized data. Wang et al.[71] developed a framework to protect sensitive attributes in high-dimensional health care data in the cloud. Also, it supports personalized privacy, and collision resistance among cloud service providers and data users. For this purpose, they implemented optimal sanitization and greedy sanitization protocols, and evaluated with real-life patient health data. The efficiency and utility of the protocols are analyzed and compared with traditional anonymization (*MulAnony*)[72] and differential privacy (*MulDiff*) approaches.[73] Li et al.[74] suggested distributed ensemble approach based privacy-preserving technique to protect privacy of patients. With the help of it, the important biomarkers are selected for making decision on type-2 diabetes. Also, the biomarkers are validated with different state patient data in U.S and compared the proposed approach with local Ad-aboost (LOCAL_Ada) and MultBoost algorithms.[75] They can extend their method with k-anonymity idea to distribute data more accurately without disclosing patients sensitive data.

Using Microaggregation

Martínez et al.[76] suggested a semantic framework to enable anonymization of structured non-numerical medical data. To manage non-numeric data, the three operators, comparison, aggregation and sorting, are presented. The framework is applied to three well-known statistical disclosure control (SDC) methods such as recoding, microaggregation and resampling, and evaluated using a real clinical dataset with structured non-numerical attributes. The result revealed that this framework produced anonymized dataset with better data utility from a semantic viewpoint. A knowledge-based numerical mapping method is offered for nominal attributes and which is useful to calculate covariance functions, coherent mean and variance for nominal data mathematically and semantically.[77] Using these functions and measures, the mapping permits adapting numerically-oriented statistical disclosure control (SDC) methods[78,79] during anonymization of nominal data. The empirical study showed that the proposed mapping preserves the semantics of raw data and produces better anonymized data for clinical research purpose. A data recipient-centered utility based de-identification framework is recommended.[51] In this, *k*-means clustering and statistical models, for example, linear regression, logistic regression and Cox's proportional hazards model, are analyzed based on the data recipient plans regarding the data. Then, a customized de-identification method is designed by enhancing condensation method[80] to satisfy recipient requirements.

This method is validated using real colon and lung cancer datasets and showed that the performance of this customized method is better when compared with other common de-identification algorithms. Table 1 shows a summary of the above discussed privacy preserving health care data publishing by different anonymization methods.

ENCRYPTION METHODS

Huang et al.[81] developed a method to safeguard patients' data in portable EHRs. Using this automatic system, patients are able to protect their data themselves by adding/deleting items. This system involves de-identification and pseudonymity, encryption, re-identification and recovery processes. The results showed that this method is feasible and effectively guarantee privacy and security of patients' data. A unified access control mechanism is suggested to reduce complexity in patient health data aggregation and privacy preservation in distributed EHRs.[82] In 2012, Rodgers et al.[83] proposed a method to evaluate the impact of the environment in individual's health by anonymously link demographic and health data. Bos et al.[38] have discussed potential practical applications of homomorphic encryption scheme to preserve the privacy of confidential health data. They showed that the functioning of a cloud service can be used to conduct predictive analysis task on homomorphically encrypted data. Then, this prediction service returns the probability of patients who are suffering from the cardiovascular disease in encrypted form. In addition, the authors proposed an automatic parameter selection module for implementing the practical homomorphic encryption. It ensures correctness and security of the results when Cox proportional hazard regression and logistic regression models employed in the predictive analysis. Thilakanathan et al.[84] have addressed privacy and security issues in the field of cloud and mobile telecare. They have developed a secure data sharing model and protocol for the cloud setting using ElGamal-based proxy re-encryption scheme.[85] They revealed that this protocol handles the user revocation problem and large data sizes, and also permits health monitoring through the Cloud. Yang et al.[86] have suggested a hybrid solution for secure sharing of medical data in a cloud setting. The cryptography method and statistical analysis are innovatively combined to provide the better balance between the privacy and data utility. The authors validated the effectiveness of their method based on the implementation of basic components such as privacy preserved query processing, integrity assurance, data merging and vertical data partition. Wang et al.[87] have proposed a fair remote retrieval model to retrieve fairly encrypted outsourced private medical records to remote untrusted cloud servers. This model achieves either the members of the research committee cooperatively disclose the actual medical records or no one of them can acquire any information in the medical records. The authors formally proved that the proposed scheme is secure under the assumption of computational Diffie-Hellman in the random oracle model.[88] In addition, the performance analysis showed that this scheme is efficient in terms of communication and computation complexity. Subsequently, Mohammed et al.[89] suggested semantically secure encryption method to ensure privacy of health data in outsourced database. Liang et al.[90] suggested two schemes, attribute-oriented authentication and transmission, to share health data through health social networks (HSN) in secure and privacy preservation manner. All the HSN user anonymize their sensitive attributes using attribute-oriented authentication scheme and then guarantee health data confidentiality while sharing by attribute-oriented transmission scheme. Therefore, these schemes protect data from various attacks effectively. Guan et al.[91] developed a privacy-preserving protocol to prevent identification of children's identities when it is transferred for analysis by internet. For this purpose, they used identity-based encryption method[92] and proved the correctness of the developed protocol. Fabian et al.[93] designed a new architecture for medical big data sharing among various organizations collaboratively and securely in semi-honest cloud settings. To provide more privacy and security for patient data, the attribute-based encryption[94] and cryptographic secret sharing methods[95] are adopted. The experimental result shows efficiency and feasibility of their approach.

Wireless Body Area Networks (WBANs)

Zhang et al.[96] suggested a priority based health data aggregation method that helps mobile users to securely forward various kinds of health data to the cloud. It avoids identity disclosure and data forgery, and reduces communication overheads in cloud assisted WBANs. A secure multi-biometric based framework is designed to protect mobile health care data with WBANs in the cloud.[97] For this, they generated random key and verified by DIEHARD testing (http://www.phy.duke.edu/~rgb/General/dieharder.php.). Then, patient's data privacy is preserved by encryption and securely stored in the cloud. Subsequently, Zhou et al.[98] proposed privacy preserving key management scheme to prevent disclosure of the patient's identity from time- and location-based attacks. It is performed by modified Blom's symmetric key technique[99] and proactive secret sharing[100] in mobile health care social networks with cloud settings. In addition, the proposed scheme's efficiency is evaluated and compared with existing schemes such as Liu's scheme[101], PSKA scheme[102], E-G scheme[103] and q-composite scheme.[104] Lin et al.[105] introduced dynamic noise threshold concept with differential privacy to preserve big sensitive data in body sensor network. They experimentally proved that this method is efficient and also protect data from attackers who known the background knowledge.

CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, we have briefly reviewed the privacy preservation methods that have been used to publish health care data. We mainly discussed how anonymization and encryption methods have been used for health care data protection in recent years and also presented their limitations. We highlight some future research directions in the disclosure of health care data publishing as follows: Privacy methods need to be suggested for complex data publishing since existing methods have some limitations. Since health dataset grows rapidly and unstructured nature of it, the efficient privacy preserving algorithms need to be offered against privacy threats. The quality of data should not be affected by privacy preserving algorithms to get the appropriate result by researchers. Also, more efficient algorithm need to be developed to preserve sensitive health data in the cloud environment.

ACKNOWLEDGEMENT

The authors acknowledge the Sathyabama University for the support.

CONFLICT OF INTEREST

Conflict of interest declared none.

REFERENCES

1. Fung BC, Wang K, Chen R, Yu PS. Privacy-preserving data publishing: A survey of recent developments. ACM Comput Surv (CSUR). 2010 Jun 1;42(4):14.
2. Tsai J, Bond G. A comparison of electronic records to paper records in mental health centers. Int J Qual Health C. 2008 Apr 1;20(2):136.
3. Kierkegaard P. Electronic health record: Wiring Europe’s healthcare. CLSR. 2011 Sep 30;27(5):503-15.
4. Qiao Y, Asan O, Montague E. Factors associated with patient trust in electronic health records used in primary care settings. Health Policy Technol. 2015 Dec 31;4(4):357-63.
5. Tamersoy A, Loukides G, Nergiz ME, Saygin Y, Malin B. Anonymization of longitudinal electronic medical records. IEEE T Inf Technol B. 2012 May;16(3):413-23.
6. Gostin LO, Levit LA, Nass SJ, editors. Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research: 2009 Feb 24; Washington: National Academies Press.
7. Sellapan P, Ng YH. A Tool for Healthcare Information Integration. JICT. 2006;5:29-44.
8. Lin YM, Zakariah MI, Mohamed A. Data leakage in ICT outsourcing: risks and countermeasures. JICT. 2010; 9: 87-109.
9. Dubovitskaya A, Urovi V, Vasirani M, Aberer K, Schumacher MI. A Cloud-Based eHealth Architecture for Privacy Preserving Data Integration. Proceedings of the ICT Systems Security and Privacy Protection; 2015 May 26; Springer; 2015. p. 585-98.
10. Malin B, Karp D, Scheuermann RH. Technical and policy approaches to balancing patient privacy and data sharing in clinical and translational research. J Invest Med. 2010 Jan 1;58(1):11-8.
11. Loukides G, Gkoulalas-Divanis A. Utility-aware anonymization of diagnosis codes. IEEE J Biomed Health Inform. 2013 Jan;17(1):60-70.
12. El Emam K, Jonker E, Arbuckle L, Malin B. A systematic review of re-identification attacks on health data. PloS One. 2011 Dec 2;6(12):e28071.
13. Kho AN, Cashy JP, Jackson KL, Pah AR, Goel S, Boehnke J, et al. Design and implementation of a privacy preserving electronic health record linkage tool in Chicago. J Amn Med Inform Assoc. 2015 Jun 23:ocv038.
14. Daglish D, Archer N. Electronic personal health record systems: A brief review of privacy, security, and architectural issues. Proceedings of the World Congress on Privacy, Security, Trust and the Management of e-Business; 2009 Aug 25; Delta Brunswick, Canada. IEEE; 2009. p. 110-20.
15. Appari A, Johnson ME. Information security and privacy in healthcare: current state of research. Int Journal Internet Enterprise Manag. 2010 Jan 1;6(4):279-314.
16. Perera G, Holbrook A, Thabane L, Foster G, Willison DJ. Views on health information sharing and privacy from primary care practices using electronic medical records. Int Journal Med Inform. 2011 Feb 28;80(2):94-101.
17. Box D, Pottas D. A model for information security compliant behaviour in the healthcare context. Procedia Technology. 2014 Dec 31;16:1462-70.
18. Arora S, Yttri J, Nilsen W. Privacy and security in mobile health (mHealth) research. Alcohol Res. 2014;36(1):143.
19. Gkoulalas-Divanis A, Loukides G, Sun J. Publishing data from electronic health records while preserving privacy: A survey of algorithms. J Biomed Inform. 2014;50:4-19.
20. Sruthi M, Rajkumar R. Securing the patients data on iot in healthcare. Int J Pharm Bio Sci. 2016 April; 7(2):168-72.
21. Ben-Assuli O. Electronic health records, adoption, quality of care, legal and privacy issues and their implementation in emergency departments. Health policy, 2015;119(3), 287-97.
22. Rajkumar R. A survey for monitoring patients healthcare information in cloud. Int J Pharm Bio Sci. 2016 April;7(2): 625- 29.
23. Eze B, Peyton L. Systematic literature review on the anonymization of high dimensional streaming datasets for health data sharing. Procedia Computer Science. 2015;63: 348-55.
24. Sweeney L. k-anonymity: a model for protecting privacy. Int J Uncertain Fuzz. 2002b;10: 557-70.
25. Bayardo RJ, Agrawal R. Data privacy through optimal k-anonymization. Proceedings of 21st International Conference on Data Engineering (ICDE’05); 2005 Apr 5; Washington, USA; IEEE Computer Society; 2005. p. 217-228.
26. Keyvanpour MR, Moradi SS. Classification and evaluation the privacy preservingdata mining techniques by using a data modification-based framework. Int J Computer Sci. Eng. 2011; 3(2):862-69.
27. Machanavajjhala A, Kifer D, Gehrke J, Venkitasubramaniam M. l-diversity: privacy beyond k-anonymity. ACM Trans Knowl Discov Data. 2007;1(1): p. 3.

28. Cynthia D, Smith A. Differential Privacy for Statistics: What we know and what we want to learn. J Privacy and Confidentiality, 2009;1(2):135-54.
29. Cynthia D, Aaron R. The algorithmic foundations of differential privacy. Theoretical Computer Science. 2014;9(3):211-407.
30. Bertino E, Ooi B, Yang Y, Deng RH. Privacy and ownership preserving of outsourced medical data. Proceedings of the 21st International Conference on Data Engineering; 2005 Apr 5; IEEE; p. 521-32.
31. Gardner J, Xiong L. An integrated framework for de-identifying unstructured medical data. Data & Knowl Eng. 2009;68:1441-51.
32. Zirra PB, Wajiga GM. Cryptographic algorithm using matrix inversion as data protection, JICT. 2011; 10: 67-83.
33. Rivest RL, Adleman L, Dertouzos ML. On data banks and privacy homomorphisms. NATO Adv Sci Inst Ser F-Com, 1978;4(11):169-80.
34. Gentry C. Fully homomorphic encryption using ideal lattices. Proceedings of the 41st Annual ACM Symposium on Theory of Computing; Bethesda, USA. 2009 May 31; ACM; 2009. p. 169-78.
35. Smart NP, Vercauteren F. Fully homomorphic encryption with relatively small key and ciphertext sizes. Proceedings of the Public Key Cryptography; 2010 May 26; Springer; 2010. p. 420-43.
36. Dijk MV, Gentry C, Halevi S, Vaikuntanathan V. Fully homomorphic encryption over the integers. Advances in Cryptology-Eurocrypt; 2010 May 30; Springer; 2010. p. 24-43.
37. Gentry C, Halevi S. Implementing gentry's fully-homomorphic encryption scheme. Advances in Cryptology (EUROCRYPT’11); 2011 May 15; Springer; 2011. p. 129-48.
38. Bos JW, Kristin L, Michael N. Private predictive analysis on encrypted medical data. J Biomed Inform, 2014;50:234-43.
39. Sweeney L. Achieving k-anonymity privacy protection using generalization and suppression. Int J Uncertain Fuzz. 2002a;10(5):571-88.
40. Lefevre K, Dewitt DJ, Ramakrishnan R. Incognito: Efficient full-domain k-anonymity. Proceedings of the ACM SIGMOD International Conference on Management of Data; 2005 Jun 14; Baltimore, USA. ACM; 2005. p. 49-60.
41. LeFevre K, Dewitt DJ, Ramakrishnan R. Mondrian multidimensional k-anonymity. IEEE International Conference on Data Engineering; 2006 Apr 3; Atlanta, Georgia. IEEE; 2006. p. 25-25.
42. Xu J, Wang W, Pei J, Wang X, Shi B, Fu AW. Utility-based anonymization using local recoding. Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2006 Aug 20; Philadelphia, USA. ACM; 2006. p. 785-90.
43. Campan A, Cooper N, Truta TM. On-the-fly generalization hierarchies for numerical attributes revisited. Secure Data Management, 2011;18-32.
44. Meyerson A, Williams R. On the complexity of optimal k-anonymity. Proceedings of the 23rd ACM IGMODSIGACT-SIGART PODS; 2004 Jun 14; Paris, France. ACM; 2004. p. 223-28.
45. Wang K, Fung BC, Yu PS. Template-based privacy preservation in classification problems. Proceedings of the 5th IEEE International Conference on Data Mining; 2005 Nov 27; Houston, Texas. IEEE; 2005.
46. Kisilevich S, Rokach L, Elovici Y, Shapira B. Efficient multidimensional suppression for k-anonymity. IEEE Trans Knowl Data Eng. 2010;22(3): 334-347.
47. Liu J, Wang K. On optimal anonymization for l+-diversity. Proceedings of the IEEE ICDE; 2010 Mar 1; IEEE; 2010b. p.213-24.
48. Abul O, Bonchi F, Nanni M. Anonymization of moving objects databases by clustering and perturbation. Inform Syst. 2010;35:884-910.
49. Domingo-Ferrer J, Torra V. A quantitative comparison of disclosure control methods for microdata. Confidentiality, Disclosure, and Data Access: Theory and practical applications for statistical agencies, 2009;111-133.
50. Batet M, Erola A, Sánchez D, Castellà-Roca J. Utility preserving query log anonymization via semantic microaggregation. Inform Sciences, 2013;242:49-63.
51. Gal TS, Tucker TC, Gangopadhyay A, Chen Z. A data recipient centered de-identification method to retain statistical attributes. J Biomed Inform. 2014;50:32-45.
52. Torra V. Microaggregation for categorical variables: A median based approach. Proceedings of privacy in statistical databases; 2004 Jun 9; Barcelona, Catalonia. Springer; 2004. p.162-74.
53. Solé M, Muntés-Mulero V, Nin J. Efficient microaggregation techniques for large numerical data volumes. Int J Inform Secur. 2012;11(4): 253-67.
54. Navarro-Arribas G, Torra V. Tree-based microaggregation for the anonymization of search logs. Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology; 2009 Sep 15; Washington, USA. IEEE; 2009. p. 155-58.
55. Loukides G, Liagouris J, Gkoulalas-Divanis A, Terrovitis M. Disassociation for electronic health record privacy. J Biomed Inform. 2014 Aug 31;50:46-61.
56. Tian H, Zhang W. Extending ℓ-diversity to generalize sensitive data. Data Knowl Eng. 2011 Jan 31;70(1):101-26.
57. Sun X, Li M, Wang H. A family of enhanced (L, α)-diversity models for privacy preserving data publishing. Future Gener Comp Sy. 2011 Mar 31;27(3):348-56.
58. Mohammed N, Fung BC, Hung PC, Lee CK. Anonymizing healthcare data: a case study on the blood transfusion service. Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining; 2009 Jun 28; Paris, France. ACM; 2009. p. 1285-94
59. Loukides G, Denny JC, Malin B. The disclosure of diagnosis codes can breach research participants'

privacy. J Am Med Inform Assoc. 2010 May 1;17(3):322-7.

60. Dwork C, McSherry F, Nissim K, Smith A. Calibrating noise to sensitivity in private data analysis. Proceedings of the Third conference on Theory of Cryptography; 2006 Mar 4; Springer; 2002. p. 265-84.

61. Chen R, Mohammed N, Fung BC, Desai BC, Xiong L. Publishing set-valued data via differential privacy. Proceedings of the VLDB Endowment; 2011 Aug 29-Sep 3; Seattle, Washington. 2011. p.1087-98.

62. Jafer Y, Matwin S, Sokolova M. Using Feature Selection to Improve the Utility of Differentially Private Data Publishing. Procedia Computer Science. 2014 Dec 31;37:511-6.

63. Newman DJ, Hettich S, Blake CL, Merz CJ. UCI repository of machine learning databases. 1998.

64. Loukides G, Gkoulalas-Divanis A, Malin B. Anonymization of electronic medical records for validating genome-wide association studies. Proceedings of the National Academy of Sciences; 2010 Apr 27; Springer; 2010. p. 7898-903.

65. Mohammed N, Jiang X, Chen R, Fung BC, Ohno-Machado L. Privacy-preserving heterogeneous health data sharing. J Am Med Inform Assoc. 2013 May 1;20(3):462-9.

66. Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman LW, Moody G, et al. Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): a public-access intensive care unit database. Crit Care Med. 2011 May;39(5):952.

67. Frank A, Asuncion A. UCI Machine Learning Repository. 2010.

68. Khokhar RH, Chen R, Fung BC, Lui SM. Quantifying the costs and benefits of privacy-preserving health data publishing. J Biomed Inform. 2014 Aug 31;50:107-21.

69. Heatherly R, Denny JC, Haines JL, Roden DM, Malin BA. Size matters: How population size influences genotype–phenotype association studies in anonymized data. J Biomed Inform. 2014 Dec 31;52:243-50.

70. Kim S, Sung MK, Chung YD. A framework to preserve the privacy of electronic health data streams. J Biomed Inform. 2014 Aug 31;50:95-106.

71. Wang W, Chen L, Zhang Q. Outsourcing high-dimensional healthcare data to cloud with personalized privacy preservation. Comput Net. 2015 Sep 9;88:136-48.

72. Liu J, Wang K. Anonymizing transaction data by integrating suppression and generalization. Proceedings of the 14th Pacific-Asia Conference on Knowledge Discovery and Data Mining; Hyderabad, India; Springer; 2010a. p. 171-80.

73. Dwork C. Differential privacy. Proceedongs of 33rd International Colloquium on Automata, Languages and Programming; Venice, Italy. Springer; 2006; p. 1-12.

74. Li Y, Bai C, Reddy CK. A distributed ensemble approach for mining healthcare data under privacy constraints. Inform Sciences. 2016 Feb 10;330:245-59.

75. Gambs S, Kégl B, Aïmeur E. Privacy-preserving boosting. Data Min Knowl Disc. 2007 Feb 1;14(1):131-70.

76. Martínez S, Sánchez D, Valls A. A semantic framework to protect the privacy of electronic health records with non-numerical attributes. J Biomed Inform. 2013 Apr 30;46(2):294-303.

77. Domingo-Ferrer J, Sánchez D, Rufian-Torrell G. Anonymization of nominal data based on semantic marginality. Inform Sciences. 2013 Sep 1;242:35-48.

78. Hundepool A, Domingo-Ferrer J, Franconi L, Giessing S, Lenz R, Naylor J, et al. Handbook on Statistical Disclosure Control (v. 1.2), Network of Excellence in the European Statistical System in the field of Statistical Disclosure Control (2010). Available on-line at http://neon. vb. cbs. nl/casc/SDC_Handbook. pdf.

79. Hundepool A, Domingo-Ferrer J, Franconi L, Giessing S, Nordholt ES, Spicer K, et al. Statistical disclosure control. John Wiley & Sons; 2012 Jul 5.

80. Aggarwal CC, Yu PS. A condensation approach to privacy preserving data mining. Proceedings of the 9th International Conference on Extending Database Technology EDBT; Heraklion, Greece. Springer; 2004. p. 183-99.

81. Huang LC, Chu HC, Lien CY, Hsiao CH, Kao T. Privacy preservation and information security protection for patients' portable electronic health records. Comput Biol Med. 2009 Sep 30;39(9):743-50.

82. Jin J, Ahn GJ, Hu H, Covington MJ, Zhang X. Patient-centric authorization framework for electronic healthcare services. Comput Secur. 2011 May 31;30(2):116-27.

83. Rodgers SE, Demmler JC, Dsilva R, Lyons RA. Protecting health data privacy while using residence-based environment and demographic data. Health Place. 2012 Mar 31;18(2):209-17.

84. Thilakanathan D, Chen S, Nepal S, Calvo R, Alem L. A platform for secure monitoring and sharing of generic health data in the Cloud. Future Gener Comp Sy. 2014 Jun 30;35:102-13.

85. Tran DH, Nguyen HL, Zha W, Ng WK. Towards security in sharing data on cloud-based social networks. Proceedings of the 8th International Conference on Information, Communications and Signal Processing (ICICS); 2011 Dec 13; IEEE; 2011. p. 1-5.

86. Yang JJ, Li JQ, Niu Y. A hybrid solution for privacy preserving medical data sharing in the cloud environment. Future Gener Comp Sy. 2015 Feb 28;43:74-86.

87. Wang H, Wu Q, Qin B, Domingo-Ferrer J. FRR: fair remote retrieval of outsourced private medical records in electronic health networks. J Biomed Inform. 2014 Aug 31;50:226-33.

88. Haralambiev K, Jager T, Kiltz E, Shoup V. Simple and efficient public-key encryption from computational diffie-hellman in the standard model. Proceedings of the Public Key Cryptography-PKC; 2010 May 26; Springer; 2010. p. 1-18.

89. Mohammed N, Barouti S, Alhadidi D, Chen R. Secure and private management of healthcare

databases for data mining. Proceedings of the 28th International Symposium on Computer-Based Medical Systems; 2015 Jun 22; IEEE; 2015. p. 191-196.

90. Liang X, Barua M, Lu R, Lin X, Shen XS. HealthShare: Achieving secure and privacy-preserving health information sharing through health social networks. Comput Commun. 2012 Sep 1;35(15):1910-20.

91. Guan S, Zhang Y, Ji Y. Privacy-Preserving Health Data Collection for Preschool Children. Comput Math Methods Med. 2013 Oct 29;2013.

92. Chatterjee S, Sarkar P. Identity-based encryption. Springer Science & Business Media; 2011 Mar 22.

93. Fabian B, Ermakova T, Junghanns P. Collaborative and secure sharing of healthcare data in multi-clouds. Inform Syst. 2015 Mar 31;48:132-50.

94. Bethencourt J, Sahai A, Waters B. Ciphertext-policy attribute-based encryption. Proceedings of the IEEE Symposium on Security and Privacy; 2007 May 20; IEEE; 2007. p. 321-34.

95. Krawczyk H. Secret sharing made short. Proceedingsof the 13th Annual International Cryptology Conference on Advances in Cryptology; 1993 Aug 22; Springer; 1994. p. 136-46.

96. Zhang K, Liang X, Baura M, Lu R, Shen XS. PHDA: A priority based health data aggregation with privacy preservation for cloud assisted WBANs. Inform Sciences. 2014 Nov 10;284:130-41.

97. Khan FA, Ali A, Abbas H, Haldar NA. A cloud-based healthcare framework for security and patients' data privacy using wireless body area networks. Procedia Computer Science. 2014 Dec 31;34:511-7.

98. Zhou J, Cao Z, Dong X, Xiong N, Vasilakos AV. 4S: A secure and privacy-preserving key management scheme for cloud-assisted wireless body area network in m-healthcare social networks. Inform Sciences. 2015 Sep 1;314:255-76.

99. Blom R. An optimal class of symmetric key generation systems. Eurocrypt 1984; 1984 Apr 9; LNCS209, Springer; 1985. p. 335-338.

100. Herzberg A, Jarecki S, Krawczyk H, Yung M. Proactive secret sharing or: how to cope with perpetual leakage. CRYPTO '95; 1995 Aug 27; LNCS963. Springer;1995. p. 339-52.

101. Liu D, Ning P. Establishing pairwise keys in distributed sensor networks. Proceedings of the 10th ACM Conference on Computer and Communication Security; ACM; 2003.

102. Venkatasubramanian KK, Banerjee A, Gupta SK. PSKA: usable and secure key agreement scheme for body area networks. IEEE T Inf Technol B. 2010 Jan;14(1):60-8.

103. Eschenauer L, Gligor V. A key management scheme for distributed sensor networks. Proceedings of the 9th ACM Conference on Computer and Communication Security; 2002 Nov 18; ACM; 2002. p. 41-7.

104. Chan H, Perrig A, Song D. Random key distribution schemes for sensor networks. IEEE Symposium on Security and Privacy; 2003 May 11; IEEE; 2003. p. 197-213.

105. Lin C, Song Z, Song H, Zhou Y, Wang Y, Wu G. Differential privacy preserving in big data analytics for connected health. J Med Syst. 2016 Apr 1;40(4):1-9.