



International Journal of Engineering, Business and Enterprise Applications (IJEBA)

www.iasir.net

DATA PRIVACY on E-HEALTH CARE SYSTEM

Abdullah Abdulrhman AlShwaier, Dr, Ahmed Zayed Emam

Department of Information Systems

King Saud University / College of Computer and Information Sciences

Saudi Arabia -Riyadh

Abstract: *The main goal of this research is to develop and implement data privacy appropriate technique that fit with E-health system. Privacy preserving data is to develop methods without increasing the risk of misuse of the data used to generate those methods. A number of effective methods for privacy preserving data mining have been proposed. Privacy preservation of sensitive information is a key factor of current era. This study concludes a comprehensive analysis against trendy privacy control. Anonymization offers more privacy options rather than other privacy preservation techniques (Randomization, Encryption, and Sanitization). However, Anonymization itself contains several techniques that require concluding best one. According to the presented analysis there is close competition among K-Anonymity, L-Diversity, T-Closeness, P-Sensitive and M-invariance. All these Anonymization methods offer resistance against prominent attacks like homogeneity and background. I conducted the analytical comparative analysis to select optimal Anonymization methods. Finally, the resulting algorithm has been implemented for both K-anonymity and L-Diversity algorithms with figures, charts and diagrams.*

Keywords: *Privacy, Privacy Preserving, Randomization, Anonymization, Encryption, K-Anonymity, L-Diversity*

I. Introduction

Rapid growth of internet technology have made possible to utilize remote communication in every aspects of life. As well as technology is exceeded, the need of privacy and security in electronic communications became hot issues. Privacy is an essential issue in case of transferring sensitive information from one location to another location through internet. Most significantly, in hospitals, Governmental organizations and industries; there is a need to publish sensitive data for purpose of performing analysis and further processing on it from other departments. The distribution of micro data is greatly reliant on data privacy and confidentiality for mining purposes. Data mining is a method of extracting useful information through mapping the proper relation among huge array of data objects [1]. Privacy control is a novel concept of data mining due to safety reasons. Naturally, humans did not want share their secrets about sensitive information with the others. Sensitive information is a bit different from secret or confidential information. Secret information associates those kinds of secrets whose loss severity is high; for example the passwords, PIN codes, credit card details etc. However, the sensitive information can be linked to that kind of information that is sensitive in nature but their disclosing severity may not be high as the secret information. The sensitive information mostly linked to diseases like HIV, Cancer, Heart problems etc. To protect this kind of information privacy preservation is important. One solution to solving this problem is to require data user to sign non-disclosure agreements. This solution will request legal resources and enforcement mechanisms. Furthermore, this cannot save against data theft even when the victim takes reasonable precautions. Thus, it is important to explore technological solutions which anonymize the data prior to its release. [1]

This situation demands to implement privacy preserving methods like (1) Randomization, (2) Anonymization, (3) Encryption, (4) Sanitization etc. in order to provide safeguard against the misuses of sensitive information. Thus this report comprehensively reviews the privacy-preserving issues, challenges and suggests the possible solutions under optimal privacy method. Each privacy technique has its own advantages and disadvantages. The potential motive behind this report is to study various kinds of privacy techniques through performing analytical comparison in order to conclude; which privacy technique grants optimal privacy. We thoroughly studied anonymization techniques like K-anonymity, L-Diversity, T-Closeness, Alpha-K-Anonymity, P-Sensitive Anatomy and M-Invariance through examples. Finally we conduct the analytical comparative analysis to select optimal anonymization method. We are mainly concerned with data custodians such as hospitals, government agencies, insurance companies, and other businesses where information required publishing publically for performing analysis or further processing. [1]

A. What is Privacy?

People frequently imagine the privacy as some sort of right. Unfortunately, the perception of a 'right' is a problematical means to begin some matter, because a right seems to be some kind of fixed standard.

There are more definition of privacy but we select the one from Wikipedia [7]. Privacy can be defined as the technique to builds the concept of 'personal space', which means free from interference by other people or a committee of peoples. [7]

Data privacy means at the time of sharing or publishing sensitive information (personal information) it should be granted that only relevant person has an access on it and no one else is able to misuse it. Personally identifiable information (PII) is a subset of confidential information that can put your identity at risk if it is lost or stolen. Another term is called trust but trust and privacy concepts are different in nature as discussed in a Table 1. [2]

Table 1: Privacy vs Trust [2]

Privacy	Trust
legal right is necessary	Access right is necessary
Only expose actual data to trustworthy people	Expose un-sensitive data to everyone.
Mostly deals with sensitive data	Deals with open (un-sensitive data) to get trust.

We can summarize the differences among Restricted **Data**, **Sensitive Data**, and Public **Data** as in the following table. The data can be classified in different ways as discussed in Table 2.

Table 2: Data Category [6]

	Restricted Data	Sensitive Data	Public Data
Legal Requirement	Protection of Data is required by law. e.g. (HIPAA)	Stanford has a contractual commitment to protect the data.	Protection of data is at the discretion of the custodian or the owner.
Access	Only those individuals designed with approved access and signed non-disclosure agreement	Stanford employees and non-employees who have a business need to know	Stanford affiliates and general public with a need to know

B. Why is privacy important?

Privacy preservation is a significant issue for data mining applications. Sensitive information disclosures can be categorized in to two kinds [3]. (1) Identity disclosure – it is happen when a specific record is linked with a specific person in a published record table. (2) Attributes Disclosure- it happens when new information is leaked against any person. Therefore before publishing any sensitive information we need privacy control mechanism. However, there are many other different reasons that people put forward to support the proposition that privacy is important. When something is private to a person, it usually means there is something within them that is considered inherently special or personally sensitive. The step to which private information is exposed therefore depends on how the public will receive this information [7].

i. Applications of Privacy

Today, privacy has variety of useful applications in several fields as discussed below.

ii. Medical

A person may not wish to share his medical records to others. They have concern that it might affect their employment or insurance coverage or maybe he would not like to involve other people in medical or psychological conditions or treatments which would be embarrassing. Revealing medical data (record) could also reveal other details about one's personal life. Privacy Breach Physicians and psychiatrists in many cultures and countries have standards for doctor-patient relationships which include maintaining confidentiality. [7]

iii. Financial

Information about a person's financial transactions, including the amount of assets, positions held in outstanding debts, stocks or funds and purchases can be sensitive. If criminals gain access to information such as a person's accounts or credit card numbers, that person could become the victim of fraud or identity theft. Information about a

person's purchases can reveal a great deal about that person's history, such as places he has visited, whom he has contacted with, products he has used, his activities and habits. [7]

iv. Political

Political privacy has been a concern since voting systems emerged in the past. The secret ballot is the simplest and most widespread measure to ensure that political views are not known to anyone other than the voter themselves—it is nearly universal in modern democracy, and considered to be a basic right of citizenship. In fact even where other rights of privacy do not exist, this type of privacy very often does. [7]

II. Related Work

The goal of privacy preserving data is to develop methods without increasing the risk of misuse of the data used to generate those methods. The topic of privacy preserving data has been extensively studied by the data community in recent years. A number of effective methods for privacy preserving data mining have been proposed. Most methods use some form of transformation on the original data in order to perform the privacy preservation. The transformed dataset is made available for mining and must meet privacy requirements without losing the benefit of mining. [10] [11]. Privacy control deals with various kinds of methods that can provide privacy against sensitive attributes. Most commonly used privacy techniques are as follows:-

1. Anonymization
2. Randomization
3. Encryption
4. Data Sanitization

A. Anonymization Methods

Anonymity is derived from the Greek word, *anonymia* (ἀνωνυμία) birthed to anonymity which means "without an identity (name)" or "namelessness". We can define anonymization as, the method that provides privacy against the sensitive information of any person by using generalization and suppression method and as a result the sensitive information cannot be distinguished (find out). Where Generalization means precise (less) the detailed information into small generalized form which can be discovered through Quasi-Identifier (QI) [14]. While generalizing any sensitive information we did not delete the complete value but we just hide its some parts or attributes. For example in Figure 2 the age of first two rows are generalized into as single digit "2" and now the digit "2" represent the both ages. Suppression means to delete the complete sensitive value and replace with (*) or any other symbol [14]. The term micro data used in anonymization means raw information without the naming identity of any person.

Anonymization method targets at making the individual record are indistinguishable among a group records by using techniques of generalization and suppression and the representative anonymization method is k-anonymity. The motivating factor behind the k-anonymity approach is that many attributes in the data can often be considered quasi-identifiers which can be used in conjunction with public records in order to uniquely identify the records. Many advanced methods have been proposed, such as, p-sensitive k-anonymity, (a, k)-anonymity, l-diversity, t-closeness, M-invariance, Personalized anonymity, and so on. [14]

i. Anonymization Techniques [13].

Anonymization methods have many techniques as below:-

1. K-Anonymity
2. L-Diversity
3. T-Closeness
4. P-Sensitive Anonymity
5. M-Invariance

ii. K-Anonymity

This method deals with the utilization of generalization and suppression upon sensitive data in which anonymized table should be indistinguishable at least level of k-anonymity (K-1). The re-discovery should be possible through Quasi-Identifier mapping. We can understand the K-Anonymity through an Example of Table 3. [13]

Table 3 : Micro Data

P-ID	Name	Age	Gender	Zip Code	Disease	City	Country
25	Sarah Omer	56	F	21400	Glysodia	Qassem	Saudi Arabia
22	Mohammed Has	32	M	21443	Ear Inect	Jouf	Saudi Arabia
20	Muna Ali	38	F	21456	Dehydrat	Jeddah	Saudi Arabia
16	Ola Sayed	20	F	21472	Dengue	Jeddah	Saudi Arabia
17	Mohammed Hus	25	M	21483	Diabetes	Jeddah	Saudi Arabia
19	Moamer	34	M	21499	Diphtheria	Jeddah	Saudi Arabia
15	Susan Saeed	34	F	36411	Crohn's	Riyadh	Saudi Arabia
14	Fateema Hassan	34	F	36426	Cretinism	Abha	Saudi Arabia
2	Ali Saleh	36	M	36426	Congesti	Riyadh	Saudi Arabia
5	Zain Ali	31	M	36432	Coronary	Riyadh	Saudi Arabia
13	Nada Saad	57	F	36461	Cowpox	Riyadh	Saudi Arabia

Example of K-Anonymity at K= 2

In below Table 4 shows that $K = 2$, it means that to point out difference between these two rows is confusing. We cannot identify which person is at 14 and which person is at 14 and these values are same but both persons are different. If you try to identify a man from a released anonymous table, but the only information you have is his birth date and gender. There are k (more than one) people meet the requirement and this is called k -Anonymity. [13]

Table 4: Example of K-Anonymity at K=2

P-ID	Age	Gender	Zip Code	Disease
25	5*	*	214**	Glysodia
22	3*	*	214**	Ear Inect
20	3*	F	214**	Dehydrat
16	2*	F	214**	Dengue
17	2*	M	214**	Diabetes
19	3*	M	214**	Diphtheria
15	3*	F	364**	Crohn's
14	3*	F	364**	Cretinism
2	3*	M	364**	Congesti
5	3*	M	364**	Coronary
13	5*	*	364**	Cowpox

In this above example Zip Code column (*) denote the generalization where Age and Gender column (*) denote the suppression.

iii. L-Diversity

L-Diversity is a privacy measure. We can define it as: - sensitive attributes should be represented with variety of distinct arrangements within each Quasi-Identifier equivalence class. It associates the limitation that it did not provide sufficient privacy against attributes disclosures. It did not save the probabilistic inference attack which means some specified attribute may be occurred repeatedly or frequently in a table that leads the adversary to judge that this kind of entity have this value. Another problem L-Diversity is that sometime some information have the distinct ratio like 1% and 99% mostly in HIV+ and HIV- disease attribute because only 1% patients have HIV+ and 99% have HIV- therefore, in this case diversity is meaningless which means distinctiveness is not well representative with variety of arrangements. It also have a problem of similarity attack which means the distinct information may have semantically (looking /Position) be similar through which important information can easily be discovered. [17]

iv. T-closeness

This privacy measure means, the allocation of sensitive attributes should be Close to the allocations of all other attributes in all tables. (i.e., the distance between the two distributions should be no more than a threshold t – (a bounded rage)). Closeness means the distribution of sensitive information should semantically seem same but should be different than the original values [18]. T-Closeness means to make the whole information public in large range. In other words to publish micro table with creating large distribution of equivalence classes in such a way at the time of observation against a single person, there are many similar equivalence classes (close classes) with distinct but look like same attributes should be there for adversary. [17] [18]

These kinds of many close classes are provided but the original data is located through Quasi-Identifier (QI).

v. P-Sensitive Anonymity

It is like Alpha-K-anonymity but the sensitive value must have P time different occurrence in equivalence classes shown in below Table in sensitivity column. In (Alpha, k)-Anonymity we just make the pairs of sensitive values but in p-sensitive anonymity we also add sensitivity level. [22]

Table 5: Example of P-Sensitive [22]

Category ID	Sensitive attribute values	Sensitivity
One	HIV, Cancer	Top Secret
Two	Phthisis, Hepatitis	Secret
Three	Heart Disease, Asthma	Less Secret
Four	Flue, Indigestion	Non Secret

vi. M-Invariance

It is new method that invokes the publication of anonymized data in multiple Versions. For this purpose it deals with signature (set of different sensitive value) and m-uniqueness (each row in anonymous version should be unique). In above table (HIV, Cancer) is a signature and each have unique signature is called m-uniqueness. Figure 7 shows the representation of signature and uniqueness. [24]

Figure 1: Example of M-Invariance [24]

ID	Age	Zip code	Disease
1	[21,22]	[12k,14k]	dyspepsia
1	[21,22]	[12k,14k]	bronchitis
2	[23,24]	[18k,25k]	flu
2	[23,24]	[18k,25k]	gastritis
3	[36,41]	[20k,27k]	flu
3	[36,41]	[20k,27k]	gastritis
4	[37,43]	[26k,35k]	dyspepsia
4	[37,43]	[26k,35k]	flu
4	[37,43]	[26k,35k]	gastritis

Table 6: Generalized Comparison of Anonymization Techniques

Anonymization Techniques	Features
K-Anonymity	▪ K- Anonymity Provides protection against identity disclosure. ▪ Sensitive information can be discovered when two rows of table have same sensitive records in equivalence class [26]. ▪ According to the authors of study [26] the K-Anonymization provides not any kind of privacy actually. ▪ It did not provide protection against attribute disclosure. ▪ Sensitive information can be discovered through Homogeneity attack. ▪ Sensitive information can be discovered through Background attack. ▪ It did not provide sufficient L-Diversity in some classes of attributes.
L- Diversity	▪ In case of L-Diversity the Sensitive information can be discover when two rows of table have same sensitive records in equivalence class [26]. ▪ Naturally some sensitive attributes contains small diversity e.g. HIV +ve and HIV –ve,

	Mostly 99 patients have HIV –Ve and only 1% patients have HIV+ve so in this situation diversity creation is a problem. This is called skewness attack and L-diversity did not handle it [27]. ▪ It did not prevent similarity attack which mean when values are semantically similar then there are some chances to discover sensitive values [27]. ▪ While creating diversity the equivalence classes become many that are difficult to handle. ▪ L- diversity did not granted full prevention against attributes privacy, because through mapping, some known attributes the sensitive attribute can be discovered [27]. ▪ Different level of diversity value has different level of privacy [27]. ▪ L- Diversity is powerful and superior to K-anonymization [26].
T-Closeness	▪ In t-closeness method has lack of attributes co-relation due to independence generalization [27]. ▪ At small value of “t” data damage chances are more [27]. ▪ Small value of “t” results high computation cost [27]. ▪ It provide defense against skewness attack [18]. ▪ It provides defense against similarity attack [18]
M-Invariance	▪ Allow republication of data in multiple versions. ▪ It is not applicable in case of attribute updates [28]. ▪ It provides attributes uniqueness
P-Sensitive Anonymity	▪ It covers the drawbacks of K-Anonymity by offering at least P different values of sensitive attributes in each record [29]. ▪ Each sensitive value is linked with at least 2 values [29]. ▪ It reduces information loss through anatomy approach [29]. ▪ It is more useful in real type of sensitive data sets [23].

III. The Comparison of general Privacy Methods

The generalized comparison of Anonymization Techniques is discussed in Table 7

Table 7: Anonymization Techniques against specific parameters.

Parameters	K-Anonymization	L-Diversity	T-Closeness	(Alpha-K) Anonymity	P-Sensitive	M-Invariance
Identity Protection	✓	✓	✓	✓	✓	✓
Attribute Protection	X	✓	✓	✓	✓	✓
Applicability of Homogeneity Attack	✓	X	X	X	X	X
Applicability of Background Attack	✓	X	X	✓	X	X
Diversity Level	X	✓	✓	X	✓	✓
Sensitive attributes publishing method	Kept confidential	Kept confidential	Kept confidential	Kept confidential	Kept confidential	Publically published
Privacy Strength	Low But Did not protect continuous publishing	Powerful But Did not protect continuous publishing	High But Did not protect continuous publishing	Low But Did not protect continuous publishing	Good But Did not protect continuous publishing	High and protect continuous publishing

IV. Selected Algorithms

A. K-Anonymization Algorithm

In this algorithm we used two techniques one is Generalization and second is Suppression. In generalization, the data will be decrypted as per System Algorithm. The generalization applied on some of the data. In Suppression, The whole data will be decrypted. It put the star instead of Data. The reason of doing Generalization and Suppression is that the details of Patients could not be visualized and no-one can extract the patient's record. [15]

The K-Anonymization Algorithm is as follows:-

Input: An integer k , and a $[k, 2k-1]$ -cover $\gamma = \{S_1, \dots, S_t\}$ of $D = \{R_1, \dots, R_t\}$.

Output: A $[k, 2k-1]$ -clustering, γ^0 , of D .

```

1: Set  $\gamma^0 = \gamma$ .
2: while  $\gamma^0$  has intersecting subset do
3:   let  $S_j, S_t \in \gamma^0$  be such that  $S_j \cap S_t \neq \emptyset$  and let  $R \in S_j \cap S_t$ .
4:   if  $|S_j| > k$  then
5:     Set  $S_j = S_j \setminus \{R\}$ .
6:   else if  $|S_t| > k$ 
7:      $S_t = S_t \setminus \{R\}$ 
8:   else  $\{|S_j| = |S_t| = k\}$ 
9:     Remove  $S_t$  and  $S_j$  from  $\gamma^0$  and replace them with  $S_j \cup S_t$ .
10:  end if
11: end while

```

Input: Table D , integer k .

$$C = F_{[k, 2k-1]} = \{S \subset D : k < |S| < 2k - 1\}$$

Output: Table $g(D)$ that satisfies k -anonymity.

```

1: Invoke algorithm 1 with  $C = F_{[k, 2k-1]}$ 
2: Convert the resulting  $[k, 2k-1]$ -cover  $\gamma$  into a  $[k, 2k-1]$ -clustering,  $\gamma^0$ , by invoking algorithm A
3: output the  $k$ -anonymization  $g(D)$  of  $D$  that responds to  $\gamma^0$ .

```

B. L-Diversity Algorithm Flow

In L-diversity, we also applied the rule of generalization and suppression but the most important fact in L-diversity is Quasi Identifier. Quasi Identifier means (QI) it will combine two attributes and applied the algorithm on new record. [15]

L-Diversity Algorithm [15]

Input: A clustering $\gamma = \{C_1, \dots, C_t\}$ of the records in the table D ; a targeted diversity parameter $\ell \geq 1$.

Output: A coarser clustering that respects ℓ -diversity.

```

1: compute  $\text{div}(C_i)$  for all  $C_i \in \gamma$ .
2: let  $C_m$  be the cluster with minimal diversity in  $\gamma$ .
3: if  $\text{div}(C_m) \geq \ell$  then
4:   output  $\gamma$  and stop.
5: end if
6: Compute cost  $(C_i, C_m)$  for all  $C_i \in \gamma \setminus \{C_m\}$ .
7: Find the cluster  $C_i \in \gamma \setminus \{C_m\}$  for which cost  $(C_i, C_m)$  is minimal.
8: Remove  $C_i$  and  $C_m$  from  $\gamma$  and add to  $\gamma$  the cluster  $C_i \cup C_m$ .
9: Go to step 2.

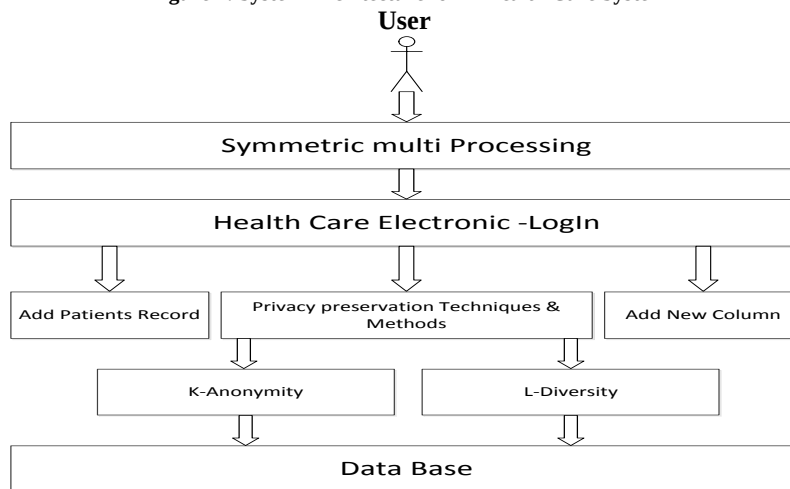
```

C. System Design

i. System Architecture for E-Health Care System

The user can access to the system from anywhere and he can enter the login ID and the Password. He can select to add, modify or delete the patient record. In the Privacy Preserving technique has an option to select or run K-Anonymity or L-Diversity with different either Generalization or Suppression and keep it in the data base

Figure 2: System Architecture for E-Health Care System



V. Result and Discussions

We are going to discuss about the Privacy preservation techniques and its benefits for developing this Algorithm. In this project, we developed two algorithms which are K-anonymity and L-diversity. This algorithm is used for hiding the identification of patients and preserving the details of patients. I represented and implemented the Algorithms with JAVA programs. We have created a large datasets for patient's records and test each technique with these data. Finally we conducted the analytical comparative analysis for these methods and this technique is very useful for today's world.

A. Result and discussions:

These are the techniques for getting secure data base for patient privacy. We select the dataset randomly from the web sites and we synthetic the data to be used during the testing.

In this Algorithm, We make the record similar so they cannot identify the actual record and we found that the L-Diversity is more secure than K-Anonymity because we will use Quasi Identifier means. We will compare between the parameters of two techniques K-Anonymity and L-Diversity and Table 8 shows the final result. These four parameters will be as the following:

1. Privacy Strength means that how the data was protected.
2. Small numbers of Datasets between 1 ~ 1000
3. Large numbers of Datasets > 1000
4. Sensitive attribute publishing that to keep the data confidential and nobody can publish it.

Table 8: Compare between K-Anonymity and L-Diversity parameters.

Parameters	K-Anonymity	L-Diversity
Privacy Strength	Low	Powerful
Small number of Dataset	Slow Response Time	Fast Response Time
Large number of Dataset	Fast Response Time	Slow Response Time
Sensitive attributes publishing method	Kept confidential	Kept confidential

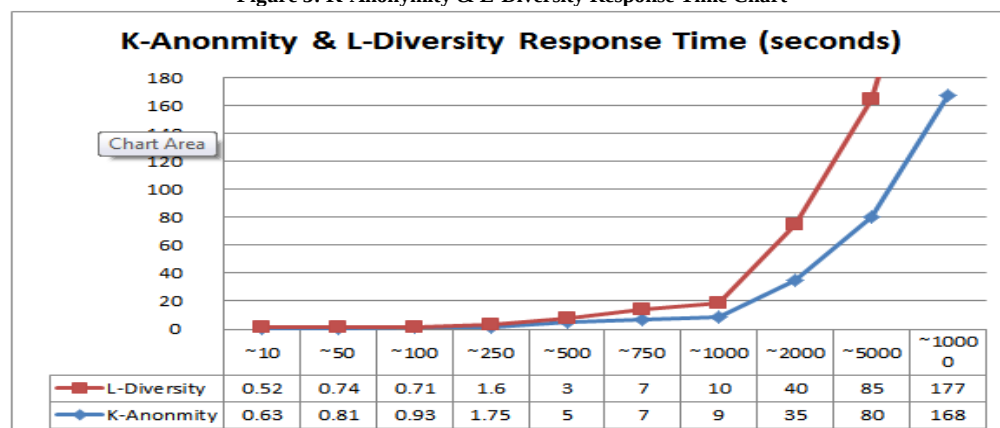
B. The Response time for K-Anonymity and L-Diversity techniques:

I used different experiment dataset for patients. I have executed the Java program and calculated the response for both K-Anonymity and L-Diversity. I selected different random Dataset 10, 50, 100, 250, 500, 750, 1000, 2000, 5000 and 10000 patients and I represented and measured the result of the time response with seconds in the following table and chart.

Table 9: Response Time between two techniques

Technique	K-Anonmity	L-Diversity
Number of records		
10	0.63 s	0.52 s
50	0.81 s	0.74 s
100	0.93 s	0.71
250	1.75 s	1.6 s
500	5.0 s	3.0 s
750	7.0 s	7.0 s
1000	9.0 s	10 s
2000	35 s	40 s
5000	80 s	85 s
10000	168 s	177 s

Figure 3: K-Anonymity & L-Diversity Response Time Chart



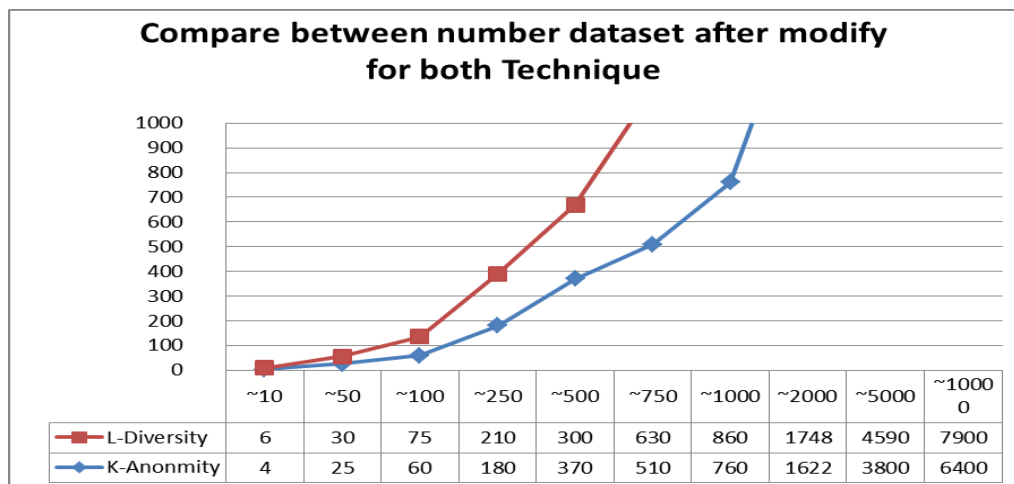
C. Compare the efficiency of implementation for K-Anonymity and L- Diversity after modify the datasets:

I represented and measured the number of dataset which are modified after implementing these K-Anonymity and L-Diversity techniques.

Table 10: Compare between numbers of modify datasets

Technique	K-Anonymity	L-Diversity
Number of records		
10	4	6
50	25	30
100	60	75
250	180	210
500	370	300
750	510	630
1000	760	860
2000	1622	1748
5000	3800	4590
10000	6400	7900

Figure 4: K-Anonymity & L-Diversity modify datasets Chart



VI. Conclusion

Privacy preservation of sensitive information is a key factor. This study concludes a comprehensive analysis against trendy privacy control as summarized in table 7. Anonymization offers more privacy options rather than other privacy preservation techniques (Randomization, Encryption, and Sanitization). However, Anonymization itself contains several techniques that require concluding the best one. I have explained and compared between different types of Anonymization. After that, I implemented two algorithms K-Anonymity and L-Diversity with JAVA programs and I used different datasets of patient's records. I represented those techniques in System Architecture, Use case diagram, Flow charts and calculated and measured the response time for each technique. Finally, I have compared between different parameters for both K-Anonymity and L-Diversity in table 8.

We are mainly concerned with data patients in medical systems at the hospitals where information required publishing publicly for performing analysis or further processing and we investigated the present status of information technology in health care area and the perceived Benefits and barriers by patient, physicians.

References

- [1] Wang P. (2010), Survey on Privacy Preserving Data Mining, International Journal of Digital Content Technology and its Applications, Volume 4, Number 9, December 2010.
- [2] Michael H, Jerome M, David J. et. Al. (2008), Resisting Structural Re-identification in Anonymized Social Networks. PVLDB08.
- [3] Li, N.; Li, T. (2007), Venkatasubramanian, S. t-Closeness: Privacy beyond k-anonymity and l-diversity. In Proceedings of IEEE 23rd International Conference on Data Engineering (ICDE07), Istanbul, Turkey, April 2007.
- [4] V.S.Verykios, E.Bertino, I.N.Fovino, L.P.Provenza, Y.Saygin, Y.Theodoridis, "State-of-the-art in Privacy Preserving Data Mining", New York, ACM SIGMOD Record, vol.33, no.2, pp.50-57,2004.
- [5] N. Zhang, "Privacy-Preserving Data Mining", Texas A&M University, pp.19-25, 2006.
- [6] Brian Markham, "Data Classification and Privacy: A foundation for compliance", University of Maryland at College Park Office of Information Technology.
- [7] <http://en.wikipedia.org/wiki/Privacy>
- [8] Office of Australian Information Commissioner - Australian Government. <http://www.privacy.gov.au/aboutprivacy/snapshot>
- [9] http://en.wikipedia.org/wiki/Privacy_Act_1988
- [10] R. Agrawal, R. Srikant, "Privacy-Preserving Data Mining", ACM SIGMOD Record, New York, vol.29, no.2, pp.439-450,2000.
- [11] Gayatri Nayak, Swagatika Devi, "A Survey on Privacy Preserving Data Mining: Approaches and Techniques", India, 2011.
- [12] Ying-hua L, Bing-ru Y. et. Al. (2011), State-of-the-art in Distributed Privacy Preserving Data Mining, 978-1-61284-486-2/11/2011 IEEE.
- [13] Wang J, Lue Y, et. Al. (2009), A Survey on Privacy Preserving Data Mining, 2009 First International Workshop on Database Technology and Applications, IEEE.
- [14] Li T. and Li N. (2007), Towards Optimal k-Anonymization, Elsevier Publisher, CERIAS and Department of Computer Science, Purdue University, 305 N. University Street, West Lafayette, IN 47907-2107, USA.

- [15] Kenig B, and Tassa T, A Practical Approximation Algorithm for Optimal k -Anonymity, Division of Computer Science, The Open University, Raanana, Israel. Cited at: http://www.openu.ac.il/Personal_sites/tamirtassa/Download/Journals/optimal_k_anon.pdf
- [16] NAYAK G. (2011), a survey on privacy preserving data mining: approaches and Techniques, International Journal of Engineering Science and Technology (IJEST), Vol. 3 No. 3, PP. (2117-2133), March 2011.
- [17] GAYATRI NAYAK, SWAGATIKA DEVI, "A SURVEY ON PRIVACY PRESERVING DATA MINING: APPROACHES AND TECHNIQUES", India, 2011.
- [18] Cao J, Karras P. et. Al. (), SABRE: A Sensitive Attribute Bucketization and Redistribution framework for t -closeness, School of Computing, National University of Singapore, cited at: <http://www.comp.nus.edu.sg/~karras/SABRE.pdf>
- [19] Gupta R. (2010), Anonymizing High Dimensional Data, CS499 Report.
- [20] Liu Y, Yang B and Li G. (2012), A Personalized Privacy Preserving Parallel (α, k) -anonymity Model, International Journal of Advancements in Computing Technology(IJACT) Volume4, Number5, PP.(265-271), March 2012.
- [21] Wong R. W, Liu Y. et. Al. (), (α, k) -anonymity Based Privacy Preservation by Lossy join, Department of Computer Science and Engineering, the Chinese University of Hong Kong. Cite at: <http://www.cse.ust.hk/~raywong/paper/waim07-alpha.pdf>
- [22] Sun X, Wang H, and Li J. et. Al. (2008), Enhanced P -Sensitive K -Anonymity Models for Privacy Preserving Data Publishing, TRANSACTIONS ON DATA PRIVACY 1 (2008), PP.(53–66).
- [23] Truta T. M, Vinay B., Privacy Protection: p -Sensitive k -Anonymity Property, Department of Computer Science, Northern Kentucky University, cited at: http://www.nku.edu/~trutat1/papers/PDM06_truta.pdf
- [24] Xiao X and Tao Y. (2008), Dynamic Anonymization: Accurate Statistical Analysis with Privacy Preservation, SIGMOD'08, June 9–12, 2008, Vancouver, BC, Canada. ACM 978-1-60558-102-6/08/06.
- [25] Leela R. P, Revathi N. (2011), Enhancing the Utility of Generalization for Privacy Preserving Re-publication of Dynamic Datasets, International Journal of Computer Applications (0975 – 8887), Volume 13– No.6, January 2011.
- [26] Mirakabad M. R. Z, Jantan A. (2008), Diversity versus Anonymity for Privacy Preservation, 978-1-4244-2328-6/08/ 2008 IEEE.
- [27] Gupta R. (2009), t -closeness: Privacy Preserving Data Mining, CS498 Report.
- [28] Zhang X, Bi H. (2010), Secure and Effective Anonymization against Re-publication of Dynamic Datasets, IEEE.
- [29] Sun X, Whang H. (2009), Achieving P -Sensitive K -Anonymity via Anatomy, 2009 IEEE International Conference on e-Business Engineering.
- [30] Wong R. C., Li J, et. Al. (2006), (α, k) -Anonymity: An Enhanced k -Anonymity Model for Privacy-Preserving Data Publishing, KDD'06, August 20–23, 2006, Philadelphia, Pennsylvania, USA