Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 5, May 2014, pg.1105 – 1110

RESEARCH ARTICLE

Implementation of Privacy Mechanism using Curve Fitting Method for Data Publishing in Health Care Domain

Varsha Meshram[1], Y.C .Bhute[2]

[1]Department of M.Tech CSE, R.T.M.N.U. Nagpur, India
[2]Asst. Prof. Department of M.Tech CSE, R.T.M.N.U. Nagpur, India
[1] varshameshram31@gmail.com; [2] yog.bhute@gmail.com

Abstract— ***In the Healthcare Domain there is a growing necessity for sharing data that contain personal data from distributed data base for Nationwide Health Information Network. To share information among hospital and other providers and supports appropriate use of Health information without taking permission of patient care with privacy protection. In this paper enlist privacy checking strategy and algorithm which developing the privacy constraints and adaptive ordering techniques for efficiently checking set of records. There are many algorithm used to provide the privacy for maintaining the database we present an alternative, so that we can publish the data with privacy. For more efficiency of database we portioned the database so the thousands of records presents in database should be present in appropriate manner. While publishing the data we make that data anonymized so that the recipient will not be able to see some sensitive information. For making the data anonymized we use the concept of trusted third party which will helpful for avoiding the potential efficacy of the database.***

Keywords— ***Data privacy, publishing data, distributed databases***

1. INTRODUCTION

We know that exchanging the information is the increasing prerequisite of the computer era. In the health care domain also it's needed to share the data for research purpose. We used the concept of data publishing for data sharing among the users who uses the health care database. The concept of data publishing is used for research and expands on the 'why, when and how' of its collection and processing, leaving an account of the analysis and conclusions to a conventional article. A data publication should include metadata describing the data in detail such as who created the data, the description of the type of data, the versioning of the data, and most importantly where the data can be accessed (if it can be accessed at all).The main purpose of a data publication is to provide adequate information about the data so that it can be reused by another researcher in the future, as well as provide a way to attribute data to its respective creator.

While publishing data [1],[3] there are some procedures are used so that sensitive data should be on safe side and no infer additional data added over there. One of the approaches is for each provider make data anonymize which leads to potential

loss of that data, we need to give individual attention. Second approach is shared data publishing [2], [4].for shared data publishing we uses the concept of trusted third party (TTP) which anonymize the data from all provider as they come from same source

We previously saw the attack which is done by external attacker, here we are introducing a new type of attack which is attack done by the internal attacker. An internal attacker attack on the sensitive data with the help of background knowledge. Which leads to the privacy break mechanism; to maintain the data on more secure side we provide many more Heurastic algorithm. Attack done by external attacker, a recipient for example P0, could be an attacker and attempts to infer additional information of the current data using the published data (T*) and some background knowledge (BK) such as easily available external data. Now considering another type of attack which is attack done by data providers using their own data; each data provider such as P1 in fig 1 can also use anonymized data T* and his own data T1 to infer additional information about other records. As compare to other attack, attack done by the internal attacker has more additional in data information of their own records, which is helpful for attack. This issue get more degraded when multiple data providers mist with each other.

T_a^*

Provider	*Name*	Age	Zip	Disease
P_1	*Alice*	[20-30]	*****	Cancer
P_1	*Emily*	[20-30]	*****	Asthma
P_3	*Sara*	[20-30]	*****	Epilepsy
P_1	*Bob*	[31-35]	*****	Asthma
P_2	*John*	[31-35]	*****	Flu
P_4	*Olga*	[31-35]	*****	Cancer
P_4	*Frank*	[31-35]	*****	Asthma
P_2	*Dorothy*	[36-40]	*****	Cancer
P_2	*Mark*	[36-40]	*****	Flu
P_3	*Cecilia*	[36-40]	*****	Flu

FIGURE 1

T_b^*

Provider	*Name*	Age	Zip	Disease
P_1	*Alice*	[20-40]	*****	Cancer
P_2	*Mark*	[20-40]	*****	Flu
P_3	*Sara*	[20-40]	*****	Epilepsy
P_1	*Emily*	[20-40]	987**	Asthma
P_2	*Dorothy*	[20-40]	987**	Cancer
P_3	*Cecilia*	[20-40]	987**	Flu
P_1	*Bob*	[20-40]	123**	Asthma
P_4	*Olga*	[20-40]	123**	Cancer
P_4	*Frank*	[20-40]	123**	Asthma
P_2	*John*	[20-40]	123**	Flu

FIGURE:2

Remaining of this paper is organized as follows: Related work for privacy measures 2. Algorithm used for maintain the privacy and its steps are explained in section 3. In Section 4 results. Finally, conclusion is discussed in section 5. The overview of proposed model is shown in Fig 3.

In Fig 3 we can saw the different module, which has different purpose. The admin Login acts as a trusted third party over here who makes authenticate to the other users. We are here defining the three curves. The first curve that has full accesses to the database. Second curve that has less access as compare to the first one and last curve that has negligible access. Patient registration module is belongs to the third curve which has negligible access.

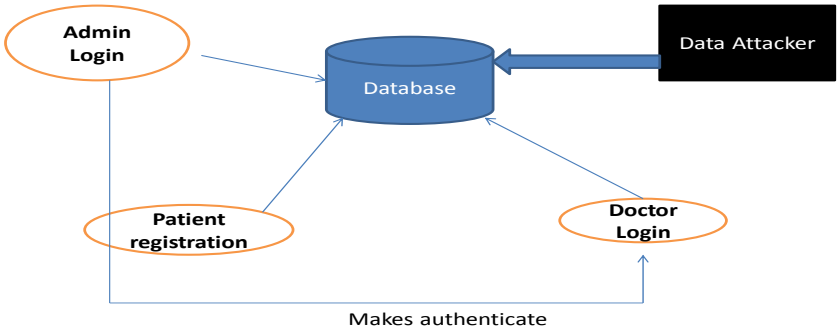


Fig.3: Overview of Proposed Model

2. RELATED WORK

Table 1 summarizes basic notations that are used to make the data anonymize. Each database consists of some data providers.

Notation	Description
T={t1,t2,t3….tn}	Set of records
DP={dp1,dp2,dp3….dpn}	Set of data providers
As	Sensitive attribute
Ds	Domain with sensitive attribute
A*	Anonymized table
C	Cross join of privacy constraints

Table 1: Notations

The standard definition of privacy checking mechanism. We are defining the privacy on the basis of K-anonymity and intersection of the l-diversity. There are two methods define to achieve the k-anonymity [10]. The first method is suppression and the second one is generalisation. Out of two we are using here the first method. Both l-diversity and the k-anonymity belong to the background knowledge of the database. To protect data from external attacker with background knowledge, we requisite some privacy constraints, C is the privacy constraint define over here.

$$C=C1\text{^}C2\text{^}\ldots\ldots\ldots Cn$$

It should satisfied the condition that C (A*) = true then we can say that A* satisfied the C.

Anonymize Mechanism

To make the data anonymize we take the help of concept l-diversity and k-anonymity. Here n are the data providers, T is the set of records and A is the anonymize mechanism. Adversary is define by I and belong to condition (m<= n-1) is an alliance of m providers, which jointly offers a set of records T1.If A*=A (T) satisfy the privacy that means it is private with respect to C.

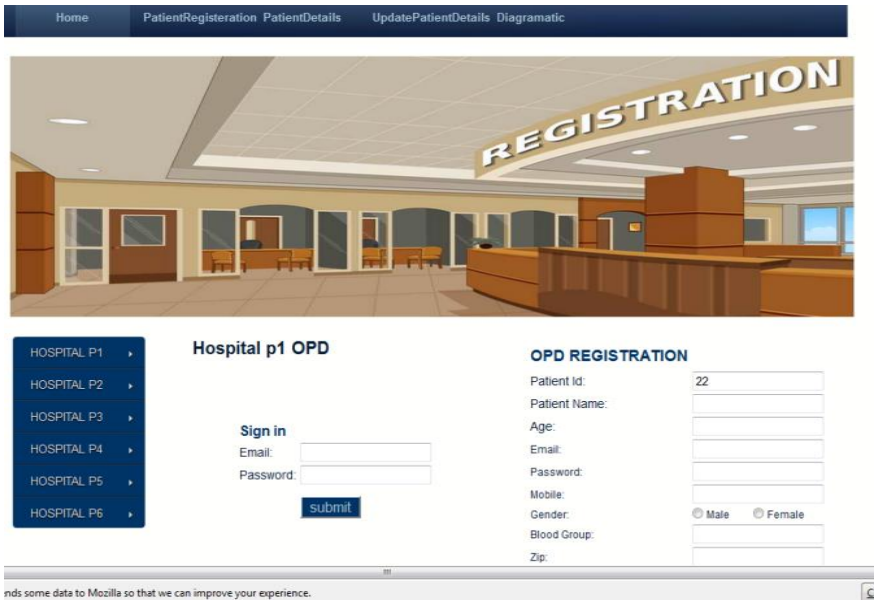
3. ALGORITHMS FOR PRIVACY

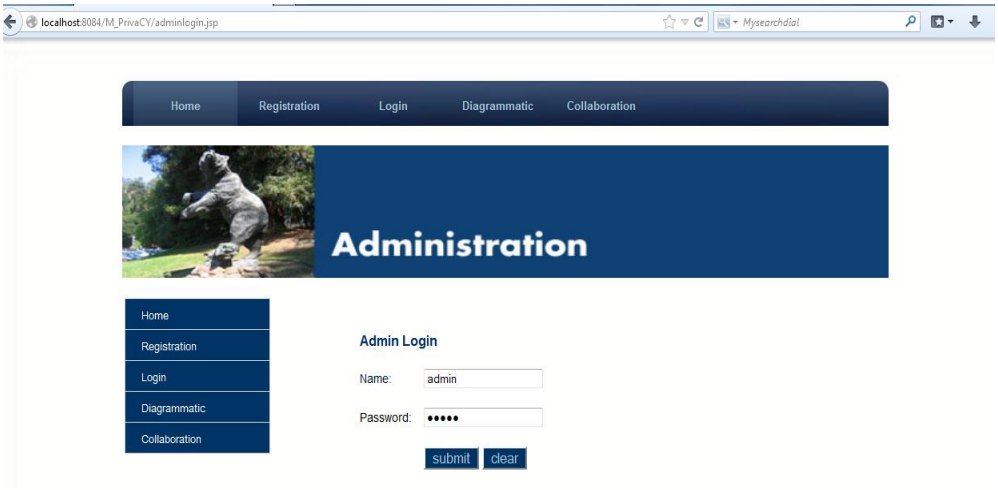
The binary verification algorithm is motivated by binary search algorithm. It checks the alliance of adversaries. The goal of each iteration is to search for a Pair Isub and Isuper. The binary verification algorithm is shown in Fig. 2.

Data: A set of records *T* provided byDP1…..DPn
Privacy constraint *C*, a privacy fitness score function F
Result: *true* if $A*$ is private, *false* otherwise
1 begin
2 sites = sort_sites(*P*, *increasing order*, F)
3 use_adaptive_order_generator(sites, *m*)
4 while is privacy_verified($A*$, *m*) = *false* do
5 *Isuper* = next_coalition_of_size($n - 1$)
6 if privacy_is_breached_by(*Isuper*) then
7 continue
8 *Isub* = next_sub-coalition_of(*Isuper;m*)
9 if privacy_is_breached_by(*Isub*) then
10 return false //early stop
11 while is_coalition_between(*Isub*, *Isuper*) do
12 *I* = next_coalition_between(*Isub*, *Isuper*)
13 if privacy_is_breached_by(*I*) then
14 *Isuper = I*
15 else
16 *Isub = I*
17 prune_all_sub-coalitions(*Isub*)
18 prune_all_super-coalitions(*Isuper*)
19 return *true*

Binary verification Algorithm

4. RESULTS





5. CONCLUSION

In this paper, we considered a new type of potential attackers in shared data publishing – an alliance of data providers. To prevent privacy disclosure by any adversary we showed that guaranteeing privacy is enough. We presented heuristic algorithms exploiting equivalence group monotonicity of privacy constraints and adaptive ordering techniques for efficiently checking privacy.

Our approach is to achieve better or comparable utility than existing algorithms while ensuring privacy efficiently. We are trying to define a proper privacy fitness score for different privacy constraints. We eliminate the unwanted data from the database so the size of database should be maintained.

REFERENCES

[1] C. Dwork, "Differential privacy: a survey of results," *in Proc. of the 5th Intl. Conf. on Theory and Applications of Models of Computation*, 2008, pp. 1–19

[2] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, "Privacy-preserving data publishing: A survey of recent Developments," *ACM Comput. Surv.* vol. 42, pp. 14:1–14:53, June 2010.

[3] C. Dwork, "A firm foundation for private data analysis," *Commun*.ACM, vol. 54, pp. 86–95, January 2011..

[4] N. Mohammed, B. C. M. Fung, P. C. K. Hung, and C. Lee, "Centralized and distributed anonymization for high-Dimensional healthcare data,"*ACMTransactions on Knowledge Discovery from Data (TKDD)*, vol.4, no. 4, pp. 18:1–18:33, October 2010.

[5] W. Jiang and C. Clifton, "Privacy-preserving distributed k-anonymity," *in Data and Applications Security XIX, ser*. Lecture Notes in Computer Science, 2005, vol. 3654, pp. 924–924.

[6] W. Jiang and C. Clifton, "A secure distributed framework for achieving k-anonymity, *VLDBJ*" vol. 15, no. 4, pp. 316–333, 2006

[7] O. Goldreich, *Foundations of Cryptography: Volume 2, Basic Applications*. Cambridge University Press, 2004.

[8] Y. Lindell and B. Pinkas, "Secure multiparty computation for privacy-preserving datamining," *TheJournal of Privacy and Confidentiality*, vol. 1, no. 1, pp. 59–98, 2009.

[9] Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity," *in ICDE*, 2006, p. 24.

[10] L. Sweeney, "k-anonymity: a model for protecting privacy," *Int. J. Uncertain. Fuzz*., vol.10, no. 5, pp. 557–570, 2002.

[11] N. Li and T. Li, "t-closeness: Privacy beyond k-anonymity and l-diversity," in *In Proc. of IEEE 23rd Intl. Conf. on Data Engineering (ICDE)*, 2007.

