INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

COMPARATIVE STUDY OF PRIVACY PRESERVING DATA PUBLISHING TECHNIQUES

MS. AMOLIKA N. PATIL[1], DR. SHRINIVAS P. DESHPANDE [2]

1. Assistant Professor, Vidyabharati Maha vidyalaya, Amravati.
2. Associate professor, HOD, DCPE HVPM, Amravati.

Accepted Date: 05/03/2015; Published Date: 01/05/2015

Abstract: In recent years, privacy is an important issue when one wants to make use of data that involves individuals sensitive information. Protection of individuals privacy is an important activity in data publishing. Data publishing generate much fear over the protection of individuals privacy. Many data holders are hesitate to provide their data for data publishing due to the fear of breaking individuals privacy. Privacy preserving data publishing (PPDP) is very important in our daily lives because these can be serious problem if the data owner releases data without protecting the sensitive information. One of the objective of privacy preserving data publishing is to hide the kind of information when data are published. In this paper various PPDP techniques are given for preserving individuals published data from the receiver.

Keywords: PPDP, Privacy, Data Publishing, K-Anonymity.

Corresponding Author: MS. AMOLIKA N. PATIL

PAPER-QR CODE

Access Online On:

www.ijpret.com

How to Cite This Article:

Amolika N. Patil, IJPRET, 2015; Volume 3 (9): 202-209

INTRODUCTION

Government and public sector websites publish huge amount of data for sharing among the departments and also to public for research. Sensitive or confidential information of individuals must be protected. Privacy is challenged through two kinds of attack that is attribute disclosure and identity disclosure [1].

A. Need of Privacy Preserving Data Publishing:

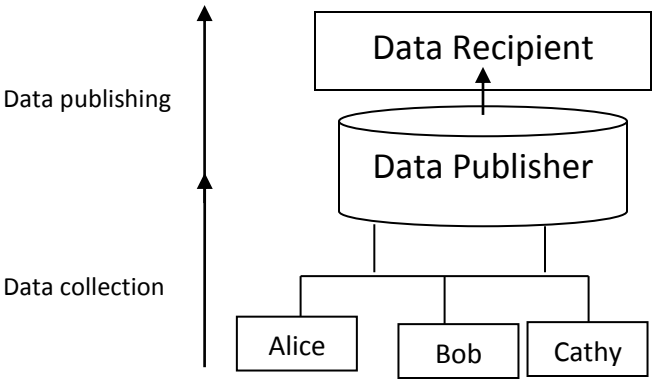
In recent years, hardware technology have lead to an increase in the capability to store and record personal data about consumers and individuals. This means that the personal data may be misused for a variety of purposes. In order to improve the Privacy of individuals data privacy preserving data publishing techniques are used[2].

Due to the rapidly increasing popularity of social networking sites on the web more and more people participate in the social networks. So any data receiver can make attack or destroy the published data which have sensitive information of the owner. Privacy is an important issue when one wants to make use of data that involves individuals' sensitive information.[3] Hence there is a need of preserving the sensitive data of the owner before published it.

B. Privacy preserving data publishing:

Today most of the organizations need to published micro data. Micro data contain records which have information about an individual entity. Means sensitive data of individuals. To prevent sensitive data from the receiver new concept is developed is called privacy preserving data publishing[4]. In privacy preserving data publishing, attributes in a table are divided into two groups: non-sensitive attributes and sensitive attributes. The values in sensitive attributes are hide from the public means it is private for individuals [5]. Privacy preserving data publishing focuses on techniques for publishing data. Privacy preserving data primarily hide the identity of record owners[6].

II DATA COLLECTION and DATA PUBLISHING PROCESS



In the data collection phase, the data holder collects data from records owners. In the data publishing phase the data holder releases the collected data to a data recipient[7].

Privacy preserving data publishing mainly focuses on designing techniques to publish data useful while preserving the privacy of individuals [8].In the basic form of privacy preserving data publishing (PPDP), the data holder has a table of the form:

D(explicit identifier, Quasi identifier, Sensitive attributes, non-sensitive attributes)

Where explicit identifier is a set of attributes, such as name and social security number, which explicitly identifies record owner. Quasi identifier is a set of attributes that potentially identifies record owners. Sensitive attributes consist of person's private or specific information and non-sensitive attributes contains all the remaining attributes which are not belongs to other categories[6][7][9].

III PRIVACY PRESERVING DATA PUBLISHING TECHNIQUES

A. Generalization:

Generalizations a data privacy method in which attributes that could cause identity disclosure are made less informative; sensitive values are replaced with a general none revealing value. An example includes replacing the gender attribute value with "person" instead of "Male" or "Female" [10][18].

In generalization technique replace the quasi identifier values with less specified values but semantically consistent. Then all quasi identifier values in a group would be generalized to the

entire group. If record in the same group must be close then to each other than generalizing the record would not lose too much information [11].

Limitations:

Generalization techniques fails on high-dimensional data due to the curse of dimensionality and it causes too much information loss due to the information distribution assumption[9][20] . By using generalization it is prove that there is big amount of data losses for high- dimensional data.

B. Bucketization:

In bucketization, tuples are divided into buckets and then to separate the sensitive attribute from the non-sensitive attributes by randomly changing the sensitive attribute values with each bucket. Bucketization is used as a method of constructing the published data from the original table. In bucketization publisher can completely mask the identifying attribute and partially mask some of the other non-sensitive attributes. While bucketization has better utility than generalization [11][12].

Limitations:

Bucketization does not prevent membership disclosure because bucketization publishes the quasi identifiers values in it's original forms. Bucketization requires a clear separation between quasi identifier and sensitive attributes. Sometimes there is a confusion that which is a quasi-identifier and which is sensitive attribute. Hence by separating the sensitive attribute from quasi identifier attribute, bucketization breaks the correlations between the Quasi identifier and sensitive attribute[9][20].

C. Suppression:

Suppression is a popular data privacy method in which data values that are unique and can be used to establish an individual's identity are omitted from the published data set [14]. Suppression technique widely used to convert a normal table into anonymous table. It is a highest level of generalization.

Limitations:

In suppression the data is suppressed and hide the original data means nothing related to actual value. But when suppression is applied on record level it causes data distortion[15][16].

D. k-anonymity:

k-anonymity is one of the popular technique of privacy preserving data publishing. The database is said to be k-anonymous when attributes are suppressed or generalized until each row is identical with at least k-1 other rows [17]. K-anonymity guarantees that the data release is accurate. K-anonymity uses two techniques: generalization and suppression [22]. The guarantee of k-anonymity is that no information can be linked to groups of less than k-individuals [9][11]. The Table is said to satisfy k-anonymity if the size of every equivalence class is greater than or equal to k[8].

Limitations:

k-anonymity technique also have some limitations such as It does not cover given individuals in the database. It opens individual's sensitive attribute. It does protect against background knowledge attack. k-anonymity algorithm fails to protect privacy[7][9][11].

E. l-Diversity:

While k-anonymity is effective in preventing identification of a record, it may not always be effective in preventing inference of the sensitive values of the attributes of that record. Therefore, the technique of l-diversity was proposed which not only maintains the minimum group size of k, but also focuses on maintaining the diversity of the sensitive attributes[19].

Limitations:

l-diversity resolves the problems of k-anonymity. But l-diversity also faced the problem that it does not prevent the probabilistic inference attack which tends to be more harmful to the human data publisher [7].

F. Slicing:

The basic idea of slicing is to overcome the limitations of generalization and bucketization. Slicing first partitions attribute into columns. Each column contains a subset or group of attribute. This is called vertically partition of the table. Slicing also partition tuples into buckets. Each bucket contains a subset of tuples. This is called horizontally partition of the table. After that each bucket value in each column are randomly change to break the associations between uncorrelated attributes [13][11]. Mainly slicing is use to break the association across columns but to prevent or preserve the association within each column.

Limitations:

Slicing preserves better utility than generalization and is more effective than bucketization for sensitive attribute. But in slicing each attribute is present in exactly one column. There is a scope of 'overlapping slicing' which releases more attribute correlations. Means sensitive attribute includes in more than one column it provide better data utility[4][21].

IV COMPARITIVE STUDY OF PRIVACY TECHNIQUES

The below TABLE. I show the different privacy models with their advantage and disadvantage

Technique	Advantages	Disadvantages
Generalization	Sensitive values are replaced with a general none revealing value.	**There is a big amount of data losses for high- dimensional data.**
Bucketization	Randomly change the sensitive attribute values with each bucket and completely mask the identifying attribute.	**Bucketization breaks the correlations between the Quasi identifier and Sensitive attribute**
Suppression	It convert a normal table into anonymous table.	**It causes data distortion.**
k-anonymity	K-Anonymity assures that the data released is perfect.	**It does not cover given individual in the database.**
l-diversity	It provides privacy preserving even when the data publisher does not know what kind of knowledge is possessed by the adversary	**Doesn't prevent the probabilistic inference attacks which tend to be more intuitive to the human data publisher.**
Slicing	**It preserves better utility and protect privacy. It can handle high dimensional data.**	**In slicing each attribute is present in exactly one column.**

V CONCLUSION

Information sharing has become part of the routine activity of many individuals, companies, organizations, and government agencies. Privacy-preserving data publishing is a promising approach to information sharing, while preserving individual privacy and protecting sensitive information. In this survey we have presented a comparative study of various privacy preserving data publishing technologies with their advantages and disadvantages. Finally it is to be concluded that the existing technologies could not fulfill the privacy preservation policy. So

the researcher is trying to develop new technology for protecting individual privacy and confidentiality in data publishing.

REFERENCES:

1. R. Mahesh, Dr. T. Meyyappam," A new method for preserving privacy in data publishing" CS & IT-CSCP, Vol 2, page no. 261-266, 19th Oct 2012.

2. V. Kavitha, M. Poornima "Disclosure Prevention in privacy preserving data publishing", IJMER Vol3, Issue 3, Pg no 1763-1767, May-June 2013.

3. Bee-Chung Chen, Daniel Kifer, Kristen LeFevre and AshwinMacnanavajjhala, "Privacy preserving data publishing" , Foundation and trends in database vol 2, page no 1-167, 2009.

4. M. Alphonsa, V. Anandam, D. Baswaraj "Methodology of privacy preserving data publishing by data slicing", IJCSMA vol.1, Issue 3, Pg no. 30-34, september 2013.

5. Bin Zhou, Jian Pei, Wo-Shun Luk, "A brief survey on anonymization techniques for privacy preserving publishing of social network data" SIGKDD Exploration Vol 10, Issue2 , 20th Aug 2007.

6. Benjamin C. M. Fung, Ke Wang, Ada Wai-Chee Fu, and Philip S. Yu, Book "Introduction to Privacy preserving data publishing concepts and techniques", CRC Press, Pg no. 1-355, Published in 2011.

7. S. Deebika, A. Sathyapriya, S.K. Kiruba "Survey result on privacy preserving techniques in data publishing" , IJLTET, Vol 3, Issue 2, Pg no 41-46 November 2013.

8. Raymond Chi-Wing Wong, Ada Wai-Chee Fu, Ke Wang and Jian Pei "Anonymization based attack in privacy preserving data publishing", ACM Transactions on Database Systems, Vol. 34, Issue 2, Article 8, June 2009.

9. Neha V. Mogre, Prof. Girish Agrawal, Prof. PragatiPatil "Privacy preserving for High-Dimensional data using anonymization technique", IJARCSSE, Vol. 3, Issue 6, Pg no 185-189, June 2013 .

10. PierangelaSamarati, Latanya Sweeney, "Protecting Privacy when Disclosing Information: k-anonymity and its enforcement through generalization and suppression" Proceedings of the IEEE Symposium on research in Security and Privacy , 1998.

11. Amar Paul Singh, Ms. DhanshriParihar "A review of privacy preserving data publishing technique", IJERMT vol. 2, Issue 6, Pg no. 32-38, June 2013.

12. David J. Martin, Daniel Kifer, Ashwin Machanavajjhala, Johannes Gehrke, Joseph Y. Halpern "Worst case background knowledge for Privacy Preserving Data Publishing" , IEEE 23rd International Conference on 15-20 April 2007.

13. Tiacheng Li, Ninghui Li, Jian Zhang, Ian Molly "Slicing: A new approach to privacy preserving data publishing", IEEE Knowledge and Data Engineering, Vol. 24, Issue 3, Pg no 561-574, March 2012 .

14. Bayardo, Agrawal ,"Data privacy through optimal k-anonymization, 21st International Conference on Engineering(ICDE'05), pg no217-228, 2005.

15. Sumit Jain, AbhishekhRaghuvanshi, "SMMCOA: Maintaining multiple correlation between overlapped attributes using slicing technique", IJETAE, Vol 3, Issue 9, Pg no 451-456, September 2013.

16. Kato Mivule, Claude Tumer "Applying data privacy techniques on tabular data in Uganda". EEE'12- The International Conference on e-learning, e-business, enterprise information system and e-goventment, las Vegas, Nevada, USA, 2011.

17. P.V.N. Prasoona, M. Vasumathi Devi, K.V. Narasimha Reddy "Privacy and utility in data publishing with full fuctional dependencies", IJETT, Vol 4, Issue 5, Pg no. 1961-1964, May 2013.

18. Wang,K., et al, 2004. Bottom-up generalization: A Data Mining Solution to Privacy Protection. In ICDM; 2004. p. 249-256

19. V.Kavitha , M.Poornima , "Disclosure Prevention in Privacy Preserving Data publishing" . IJMER Vol 3, issue 3, May-June 2013, pp 1763-1767.

20. K. Vani, B.Srinivas "Enhanced slicing for privacy preserving data publishing", IJESVol 2, Issue 10, pages 01-04, 2013.

21. Alphonsa Vedangi, V. Anandam "Data slicing technique to privacy preserving and data publishing", IJRET, Vol 02, Issue 10, Pg no 120-126, Oct 2013.

22. LuoYongcheng, Le Jiajin, Wang Jian, "Survey of Anonymity Techniques for Privacy Preserving", ISCCC 2009, Proc .of CSIT vol.1 , Pg no 248-252, (2011) IACSIT Press, Singapore.