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. 11, November 2014, pg.156 – 163

RESEARCH ARTICLE

A Novel Technique for Privacy Preserving Data Publishing

Pargi Sridhar Reddy[1], Ch.Ravi[1]

PG Scholar in the Department of Computer Science and Engineering TKR College of Engineering and Technology, Hyderabad, Telangana-500097

sridharreddy545@gmail.com

Associate Professor Department of Computer Science and Engineering TKR College of Engineering and Technology, Hyderabad, Telangana-500097

Abstract

Data Anonymization is technique used for organizing of data. Privacy Preserving Data Mining that is the study of data mining lot of Problems on privacy, it given growing thoughtfulness from the investigation community. Privacy-preservation data publishing has received lot of thoughtfulness, as it is always a problem of how to protect database of high dimension. In much organization where large number of personal data is available, such data must be protected. The personal data may be misused, for a variety of resolutions. In order to improve these apprehensions, a number of techniques have recently been proposed in order to perform the data mining tasks in a privacy-preserving way. There are several anonymization techniques available such as generalization and bucketization that are designed for privacy Preservation of micro data publishing. But it has been seen that for high dimension data generalization misses the information, bucketization on other hand does not prevent membership discovery. Here we are implementing additional anonymization technique known as Slicing. The consequence of using slicing is that it can switch high dimension data. Slicing conserves better data service than generalization and also prevents participation revelation. Here we focus on operational method that can be used for providing better data effectiveness and can handle high-dimensional data.

Keywords: *privacy, Generalization, Bucketization, Tuples, Data Publishing*

I. INTRODUCTION

There are various situations in which a person might choose to withhold their identity. Acts of charity have been performed anonymously when benefactors do not wish to be acknowledged. A person who feels threatened might attempt to mitigate that threat through anonymity. In certain situations, it is illegal to remain anonymous. In the United States, 24 states have “Stop and identify” statutes that requires persons Detained to self-identify when requested by a law enforcement officer. In recent years, due to increase in ability to store personal data about users and the increasing sophistication of data mining algorithms to leverage this information the problem of privacy-preserving data mining has become more important. A number

of anonymization techniques have been investigated in order to perform privacy-preserving data mining. Data anonymization technique for privacy-preserving data publishing has received a lot of attention in recent years. Detailed data (also called as micro data) contains information about a person, a household or an association. Most popular anonymization techniques are *Generalization and Bucketization.* There are number of attributes in each record which can be categorized as 1) *Identifiers* such as *Name or Social Security Number* are the attributes that can be uniquely identify the individuals. 2) some attributes may be Sensitive Attributes(SAs) such as *disease* and *salary* and 3) some may be Quasi-Identifiers (QI) such as *zip code*, *age*, and *sex* whose values, when taken together, can potentially identify an individual.

II. Related Work

Anonymity is the condition of having one's name or identity unknown or concealed. It serves valuable social resolutions and empowers individuals as against institutions by limiting surveillance, but it is also used by wrong doers to hide their actions or avoid accountability the ability to allow anonymous access to services, which avoid tracking of user's personal information and user behavior such as user location, frequency of a service usage, and so on. If someone sends a file, there may be information on the file that leaves a trail to the sender. The sender's information may be traced from the data logged after the file is sent.

2.1. Anonymity vs. security

Anonymity is a very powerful technique for protecting privacy. The decentralized and stateless design of the Internet is particularly suitable for anonymous behavior. Although anonymous actions can ensure privacy, they should not be used as the sole means for ensuring privacy as they also allow for harmful activities, such as spamming, slander, and harmful attacks without fear of reprisal. Security dictates that one should be able to detect and catch individuals conducting illegal behavior, such as hacking, conspiring for terrorist acts, and conducting fraud. Legitimate needs for privacy should be allowed, but the ability to conduct harmful anonymous behavior without responsibility and repercussions in the name of privacy should not.

2.2 Anonymity vs. Privacy

Privacy and anonymity are not the same. The distinction between privacy and anonymity is clearly seen in an information technology context. Privacy corresponds to being able to send an encrypted e-mail to another recipient. Anonymity corresponds to being able to send the contents of the e-mail in plain, easily readable form but without any information that enables a reader of the message to identify the person who wrote it. Privacy is important when the contents of a message are at issue, whereas anonymity is important when the identity of the author of a message is at issue.

III. Projected Work

Problem Statement

Database privacy is a concept that is important to associations and private citizens alike. Privacy professionals also can protect storage systems against theft involving servers, hard drives, desktops and laptops. Associations should ensure that storage management interfaces and all database backups, whether on-site or off-site, maintain their integrity. If attacks on a database occur, it is an association's responsibility to take defensive measures. This might first entail the immediate classification of data according to importance. Then, encryption methods might be employed to help protect applications and data based on their sensitivity levels. Of course, the best method of protecting a database's privacy is prevention. One method of database privacy protection might include assessing a database regularly for exploits and signs that it has been compromised. If an association can detect exploits or indications of database compromising before the threat becomes real and unmanageable, the database might be able to be rectified with little and reversible damage.

Objectives

An important investigate problem is for handling high-dimensional data. As per the above, Privacy Preservation for high dimensional database is important. There are two popular data anonymization technique Generalization and Bucketization. These techniques are designed for privacy preserving micro data publishing. Our Proposed work includes a slicing technique which is better than generalization and bucketization for the high dimension data sets. Slicing preserves better data utility than generalization and can be used for participation disclosure protection.

IV. SLICING ALGORITHMS

Slicing first partitions attributes into columns. Each column contains a subset of attributes. This vertically partitions the table. Slicing also partition tuples into buckets. Each bucket contains a subset of tuples.

4.1 Attribute Partition and Columns

An attribute partition consists of several subsets of A, such that each attribute belongs to exactly one subset. Each subset ofS attributes is called a column. Specifically, let there be columnsC1, C2. . . Cc, then $\cup_{i=1}^{c} C_i = A$ and for any $1 \leq i_1 \neq i_2 \leq c$, $C_{i1} \cap C_{i2} = \emptyset$. For simplicity of discussion, consider only one sensitive attribute S. If the data contain multiple sensitive attributes, one can either consider them separately or consider their joint distribution [25]. Exactly one of the c columns contains S. Without loss of generality, let the column that contains S be the last column Cc. This column is also called the *sensitive column*. All other columns {C1, C2,…..Cc-1} contain only QI attributes.

Our algorithm partitions attributes so that highly correlated attributes are in the same column. This is good for both utility and privacy. In terms of data utility, grouping highly correlated attributes preserves the correlations among those attributes. In terms of privacy, the association of uncorrelated attributes presents higher identification risks than the association of highly correlated attributes because the association of uncorrelated attributes values is much less frequent and thus more identifiable. Therefore, it is better to break the associations between uncorrelated attributes, in order to protect privacy. In this phase, first compute the correlations between pairs of attributes and then cluster attributes based on their correlations.

4.2Measures of Correlation

Two widely used measures of association are Pearson correlation coefficient [5] and mean square contingency coefficient [5]. Pearson correlation coefficient is used for measuring correlations between two continuous attributes while mean-square contingency coefficient is a chi-square measure of correlation between two categorical attributes. Choose to use the mean-square contingency coefficient because most of our attributes are categorical. Given two attributes A1 and A2 with domains fv11; v12; . . . ; v1d1g and fv21; v22; . . . ; v2d2g, respectively. Their domain sizes are thus d1 and d2, respectively.
The mean-square contingency coefficient between A1 and A2 is defined as

$$\phi^2(A_1, A_2) = \frac{1}{\min\{d_1, d_2\} - 1} \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} \frac{(f_{ij} - f_i f_j)^2}{f_i f_j}.$$

Here, f_i and f_j are the fraction of occurrences of v_{1i} andv_{2j} in the data, respectively. , f_i and f_j is the fraction of co-occurrences of v_{1i} and v_{2j} in the data.

4.3 Column Generalization

In the second phase, tuples are generalized to satisfy some minimal frequency requirement. We want to point out that column generalization is not an indispensable phase in our algorithm. As shown by Xiao and Tao [34], bucketization provides the same level of privacy protection as generalization, with respect to attribute disclosure. Although column generalization is not a required phase, it can be useful in several aspects. First, column generalization may be required for identity/ participation disclosure protection. If a column value is unique in a column (i.e., the column value appears only once in the column), a tuple with this unique column value can only have one matching bucket. This is not good for privacy protection, as in the case of generalization/bucketization where each tuple can belong to only one equivalence-class/bucket. The main problem is that this unique column value can be identifying. In this case, it would be useful to apply column generalization to ensure that each column value appears with at least some frequency.
Second, when column generalization is applied, to achieve the same level of privacy against attribute disclosure, bucket sizes can be smaller. While column generalization may result in information loss, smaller bucket-sizes allow better data utility. Therefore, there is a trade-off between column generalization and tuple partitioning. The trade-off between column generalization and tuple partitioning is the subject of future work. Existing anonymization algorithms can be used for column generalization, e.g., Mondrian . The algorithms can be applied on the sub table containing only attributes in one column to ensure the anonymity requirement.

4.4. Tuple Partitioning

In the tuple partitioning phase, tuples are partitioned into buckets, no generalization is applied to the tuples. Fig. 1 gives the description of the tuple-partition algorithm. The algorithm maintains two data structures:
1) a queue of buckets Q

2) a set of sliced buckets SB.
Initially, Q contains only one bucket which includes all tuples and SB is empty. For each iteration, the algorithm removes a bucket from Q and splits the bucket into two buckets .If the sliced table after the split satisfies l-diversity, then the algorithm puts the two buckets at the end of the queue Q.Otherwise, we cannot split the bucket anymore and the algorithm puts the bucket into SB . When Q becomes empty, we have computed the sliced table. The set of sliced buckets is SB .The main partof the tuple-partition algorithm is to check whether a sliced table satisfies '*l*-diversity .

Algorithm tuple-partition (T, l)

1. Q = {T}; SB = ∅.
2. while Q is not empty
3. Remove the first Bucket B from Q; Q= Q -{B}.
4. Split B into two Buckets B1 and B2, as in Mondrian.
5. if **diversity-check** (T,Q ∪{B1,B2} ∪ SB, **l**)
6. Q= Q ∪ {B1,B2}.
7. else SB= SB ∪{B}.
8. return SB.

Algorithm1. The tuple-partition algorithm.

Fig. 2 gives a description of the diversity-check algorithm. For each tuple t, the algorithm maintains a list of statistics L[t] about t's matching buckets. Each element in the list L[t]contains statistics about one matching bucket B: the matching probability p(t,B) and the distribution of candidate sensitive values D(t,B)

Algorithm diversity-check(T,T*,/)

1. for each tuple t ∈ T, L[t] = ∅.
2. for each bucket B in T*
3. record f(v) for each value v in bucket B.
4. for each tuple t ∈ T
5. calculate p(t,B) and find (t,B).
6. L[t]=L[t] ∪ {(p(t,B), D(t,B))}.
7. for each tuple t ∈ T
8. calculate p(t,s) for each based on L[t].
9. if p(t,s) >1// return false
10. return true.

Algorithm2 Algorithm diversity-check

V. Experimental Results

An important investigate problem is for handling high-dimensional data. As per the above, Privacy Preservation for high dimensional database is important. There are two popular data anonymization technique Generalization and Bucketization. These techniques are designed for privacy preserving micro data publishing. Our Proposed work includes a slicing technique which is better than generalization and bucketization for the high dimension data sets. Slicing preserves better data utility than generalization and can be used for participation disclosure protection. Existing data anonymization techniques can be classified in several dimensions.

1) Nature of data

Techniques have been proposed for (a) tabular data, which represents information about entities (e.g., people), their quasi-identifiers (e.g., age, gender, zip code), and their sensitive information (e.g. salary, disease); (b) item set data, which represents transactional (or "market basket") data, associating people with the sets of items purchased in a transaction; and(c) graph data, which represents sensitive associations between entities (e.g., people in social networks).

2) Anonymization approaches

Proposed anonymization techniques use a variety of approaches, including
(a) Suppression, where information is removed from the data
(b) Generalization, where information is coarsened into sets

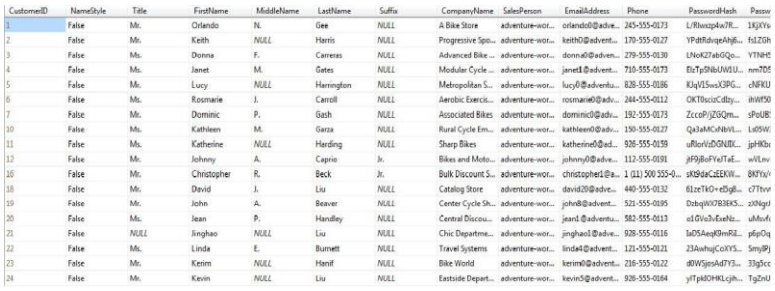
(c) Perturbation, where noise is added to the data ; and
(d) Permutation, where sensitive associations between entities are swapped.

3) Anonymization objectives

Various privacy goals are achieved by ensuring the published data has certain properties, such as (a) k-anonymity, where each individual in the database must be indistinguishable from k−1 others; (b) *l*-diversity, which seeks to ensure sufficient diversity in the sensitive information associated with individuals; and (c) other goals which aim to prevent certain inferences based on assumptions about knowledge held by an attacker.

The person must be aware of various data anonymization technique. Futher, they should also be aware of relational database, and how better privacy can be given to those records that are available in database tables.

Privacy for the database is becoming a huge problem. In many government and private associations, in Hospitals, various multinational companies, colleges etc. where there is large number of database available, privacy for such database should be maintained properly.Our implementation includes a large database of company known as Adventures Works. Its database includes tables such as Address table, Customer table, Customer Address table, Product, Product Description, Product Model table etc.



CustomerID	NameStyle	Title	FirstName	MiddleName	LastName	Suffix	CompanyName	SalesPerson	EmailAddress	Phone	PasswordHash	Passw
1	False	Mr.	Orlando	N.	Gee	NULL	A Bike Store	adventure-wor...	orlando0@adve...	245-555-0173	L/Rlwxzp4w7R...	1KjXYs
2	False	Mr.	Keith	NULL	Harris	NULL	Progressive Spo...	adventure-wor...	keith0@advent...	170-555-0127	YPdtRdvqeAhj6...	fs1ZGh
3	False	Ms.	Donna	F.	Carreras	NULL	Advanced Bike ...	adventure-wor...	donna0@adven...	279-555-0130	LNoK27abGQo...	YTNH5
4	False	Ms.	Janet	M.	Gates	NULL	Modular Cycle ...	adventure-wor...	janet1@advent...	710-555-0173	ElcTpSNbUW1U...	nm7D5
5	False	Mr.	Lucy	NULL	Harrington	NULL	Metropolitan S...	adventure-wor...	lucy0@adventu...	828-555-0186	KJqV15wsX3PG...	cNFKU
6	False	Ms.	Rosmarie	J.	Carroll	NULL	Aerobic Exercis...	adventure-wor...	rosmarie0@adv...	244-555-0112	OKT0sciiCdby...	ihWf50
7	False	Mr.	Dominic	P.	Gash	NULL	Associated Bikes	adventure-wor...	dominic0@adv...	192-555-0173	ZccoP/jZQQm...	sPoUB
10	False	Ms.	Kathleen	M.	Garza	NULL	Rural Cycle Em...	adventure-wor...	kathleen0@adv...	150-555-0127	Qa3aMCvNhVrL...	Ld5W
11	False	Ms.	Katherine	NULL	Harding	NULL	Sharp Bikes	adventure-wor...	katherine0@ad...	926-555-0159	uRlorVzDGNJBX...	jpHKbs
12	False	Mr.	Johnny	A.	Caprio	Jr.	Bikes and Moto...	adventure-wor...	johnny0@adve...	112-555-0191	jtF9jBoFYeJTaE...	wVLnv
16	False	Mr.	Christopher	R.	Beck	Jr.	Bulk Discount S...	adventure-wor...	christopher1@a...	1 (11) 500 555-0...	sK9dkCzEEKW...	8KfXv
18	False	Mr.	David	J.	Liu	NULL	Catalog Store	adventure-wor...	david20@adve...	440-555-0132	61zeTkO+d5gB...	c7Ttvv
19	False	Mr.	John	A.	Beaver	NULL	Center Cycle Sh...	adventure-wor...	john8@advent...	521-555-0195	DzbqWX7B3EkS...	zXNgrJ
20	False	Ms.	Jean	P.	Handley	NULL	Central Discou...	adventure-wor...	jean1@adventu...	582-555-0113	o1GVo3vExeNs...	uMovf
21	False	NULL	Jinghao	NULL	Liu	NULL	Chic Departme...	adventure-wor...	jinghao1@adve...	928-555-0116	laDSAeqK9mRil...	p6pOq
22	False	Ms.	Linda	E.	Burnett	NULL	Travel Systems	adventure-wor...	linda4@advent...	121-555-0121	23AwhujCoXYS...	5myBPj
23	False	Mr.	Kerim	NULL	Hanif	NULL	Bike World	adventure-wor...	kerim0@advent...	216-555-0122	d0WSjpsAd7Y3...	33g5cc
24	False	Mr.	Kevin	NULL	Liu	NULL	Eastside Depart...	adventure-wor...	kevin5@advent...	926-555-0164	yITpkIOHKLcjlh...	TgZnU

Fig. 3(a): Figure shows the Table that containsDatabase of Customer Details

Now, for such database we are providing privacy, that no customer or product information gets loss.

a) As we are slicing the database, so our database should be of very large size.

b) After proper installation, the records in database table are to be protected. First, the original table will not be shown to outer world. The original table will be with administrator, and the sliced data in which the field in the record get clubbed with some other field record, will be shown to outer world. This clubbing is based on Gold Code interleaved sequence. For this, check whether one of the Original table records are clubbed or not, so that Database security or privacy of Database is maintained.

(i) We first open the project which is named as Slicing Databases in Microsoft visual studio2010.

(ii) At the time of execution, firstly the tables of the database are loaded.

(iii) For slicing, **apply slicing ()** class is created, by which the data are sliced.

(iv) A original DB Address table is present. Two slice of this original table has been created, one is **DB Address slice 1** and another is **DB Address slice2.** Class Database Operation is created, and all the records are read by **(sqlReader.Read),** then in try catch all the fields of the database are entered like "Address ID", "AddressLine1", "AddressLine2", "City", "State Province" , "Country code" etc. All this information is added to slice table. In slice 1, DB Address slice1class is created, in which all the fields of that are present in slice 1 table, is entered. Like, "Address ID", "AddressLine1", "AddressLine2" , "City" are in slice 1. Now, for slice 2 another class is created DB Address slice 2 in which remaining fields of table are present. e.g. "State Province", "Country code" etc. In this way, original table field is sliced.

(v) Secondly, database table are sliced by using **Gold Code Random Interleaved Sequence Algorithm.** Which 'Generate a slicing pattern. Random class is generated, this class represents pseudo random number generator, i.e. a device that produce a sequence of number that meet certain statistical requirement of randomness.

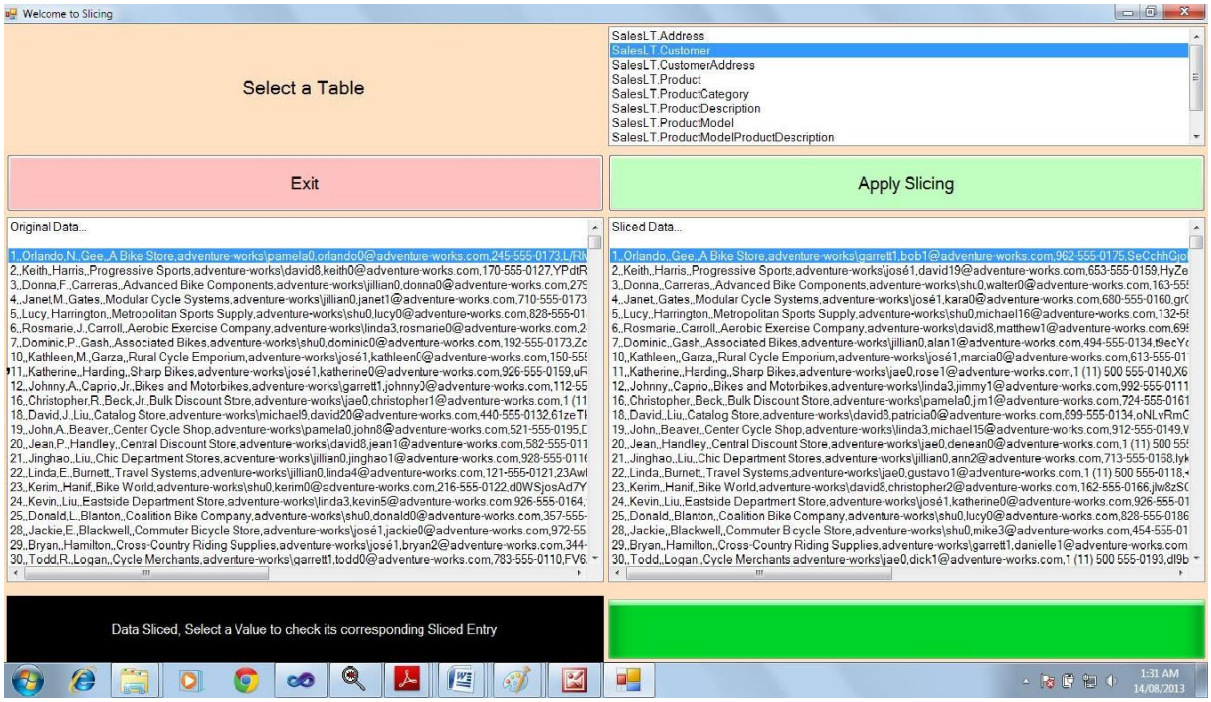


Fig. 3(b): Figure shows the Slicing applied on Database of Customer Details

Fig. 3(a) and Fig. 3(b) describe the database table of Customer details and apply slicing technique on it. This table consist large number of columns such as Customer ID, Name, Address, Email id, Phone Number etc. such a large tables which is having large number of columns and fields in it i. e. high dimensional database, such database with personal information must be maintained with some privacy technique. Our fields in original database is saved with protectly and to outer world the sliced database is shown. Fig.3(a) shows such high dimensional database and in Fig 3(b) it shows that on left side there is a Original Data and on right side there is a Sliced Data in first row four fields are same in original and sliced data but it gets changed after fourth field it shows "**Pamelo0**" in right side and "**garrett1**" on left side i.e. Sliced Data, fifth field is "**orlando0@adventure-works.com**" on Original data and same field is sliced with "**bob1@adventure-works.com**".
Likewise, many of the fields are changed by applying slicing technique.

The basic steps that are included are:

a. Generate a pattern number
b. Check if this number is already present in the pattern
c. if present then create new number
d. Add this to the list
e. Slicing Pattern Generated!
f. Slice 1 is from normal count, and Slice 2 is from Generated Pattern Count
g. Now show both the data's at the output

VI. Conclusion

The anonymization, can provide strong and robust privacy protection to individuals in published or shared databases without sacrificing much utility of the data. Anonymity is very powerful technique for protecting privacy. This paper presents a new approach for privacy preservation called Slicing. Slicing is promising technique for handling high-dimensional data. By using slicing for the large datasets, can help to hide the original data from real world, identity the records will be changed or removed and then shown to the real world. This makes database more protect and also keep data privacy. Our comparison proves that slicing is better than generalization and Bucketization. In comparison it has been shown that, for high dimension data generalization loses considerable amount of information. And Bucketization does not prevent participation disclosure.

It has been proved that Slicing preserves the datasets of any large size. It is flexible for any large database i.e. it is better than other previous technique generalization and Bucketization. Anonymity in Database can be maintained properly.

Here we discussed on Three Dimensions

(1) Designing a simple, intuitive, and robust privacy model
(2) Designing an effective anonymization technique that works with real-world databases
(3) Developing a framework for evaluating privacy and utility tradeoff.

References

[1] Tiancheng Li, Ninghui Li, Senior Member, IEEE, Jia Zhang, Member, IEEE, and Ian Molloy "Slicing: A New Approach for Privacy Preserving Data Publishing" *Proc. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 3, MARCH 2012.*

[2] V. Ciriani, S. De Capitani di Vimercati, S. Foresti, and P. Samarati On K-Anonymity. In Springer US, Advances in Information Security (2007).

[3] Latanya Sweeney. k-anonymity: "a model for protecting privacy". *International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems, 10(5):557–570, 2002.*

[4] J. Brickell and V. Shmatikov, "The Cost of Privacy: Destruction of Data-Mining Utility in Anonymized Data Publishing," *Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), pp. 70-78, 2008*

[5] Benjamin C. M. Fung, Ke Wang, Ada Wai-Chee Fu, and Philip S. Yu, "Privacy Preserving Data Publishing Concepts and Techniques" ,*Data mining and knowledge discovery series (2010).*

[6] Neha V. Mogre, Girish Agarwal, Pragati Patil: "A Review on Data Anonymization Technique For Data Publishing" *Proc. International Journal of Engineering Investigate & Technology (IJERT) Vol. 1 Issue 10, December- 2012 ISSN: 2278-0181*

[7] N. Li, T. Li, and S. Venkatasubramanian, "t-Closeness: Privacy Beyond k-Anonymity and '*l*-Diversity," *Proc. IEEE 23rd Int'l Conf. Data Eng. (ICDE), pp. 106-115, 2007.*

[8] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. "*l*-diversity: Privacy beyond k-anonymity". *In ICDE, 2006.*

[9] D. Martin, D. Kifer, A. Machanavajjhala, J. Gehrke, and J. Halpern. "Worst-case background knowledge for privacy-preserving data publishing". *In ICDE, 2007.*

[10] G.Ghinita, Y. Tao, and P. Kalnis, "On the Anonymization of Sparse High-Dimensional Data," *Proc. IEEE 24th Int'l Conf. Data Eng. (ICDE), pp. 715-724, 2008.*

[11] R. J. Bayardo and R. Agrawal, "Data Privacy through Optimal k- Anonymization," in *Proc. of ICDE*, 2005, pp. 217–228.

[12] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan, "Incognito: Efficient Full-domain k-Anonymity," in *Proc. of ACM SIGMOD*, 2005, pp. 49– 60.

[13] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan, "Mondrian Multidimensional k-Anonymity," in *Proc. of ICDE*, 2006.

[14] Gabriel Ghinita, Member IEEE, PanosKalnis, Yufei Tao," Anonymous Publication of Sensitive Transactional Data" in *Proc. of IEEE Transactions on Knowledge and Data Engineering* February 2011 (vol. 23 no. 2) pp. 161-174.

[15] D.J. Martin, D. Kifer, A. Machanavajjhala, J. Gehrke, and J.Y. Halpern, "Worst-Case Background Knowledge for Privacy-Preserving Data Publishing," *Proc. IEEE 23rd Int'l Conf. Data Eng. (ICDE), pp. 126-135, 2007*.

[16] X. Xiao and Y. Tao, "Anatomy: Simple and Effective Privacy Preservation*," Proc. Int'l Conf. Very Large Data Bases (VLDB), pp. 139-150, 2006.*

[17] Y. He and J. Naughton, "Anonymization of Set-Valued Data via Top-Down, Local Generalization," *Proc. Int'l Conf. Very Large Data Bases (VLDB), pp. 934-945, 2009.*

[18] D. Kifer and J. Gehrke, "Injecting Utility into Anonymized Data Sets," *Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), pp. 217-228, 2006.*

[19] T. Li and N. Li, "On the Tradeoff between Privacy and Utility in Data Publishing," *Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), pp. 517-526, 2009.*

[20] Y. Xu, K. Wang, A.W.-C. Fu, and P.S. Yu, "Anonym zing Transaction Databases for Publication," *Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), pp. 767-775, 2008.*

[21] J. Xu, W. Wang, J. Pei, X. Wang, B. Shi, and A.W.-C. Fu, "Utility- Based Anonymization Using Local Recoding," *Proc. 12th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), pp. 785-790, 2006.*

[22] Neha V. Mogre, Girish Agarwal, Pragati Patil "Privacy Preserving for High-dimensional Data using Anonymization Technique " *Proc. International Journal of Advanced Investigate in Computer Science And Software Engineering(IJARCSSE) Vol. 3 Issue 6, JUNE- 2013 ISSN: 2277-128X*

[23] G. T. Duncan and D. Lambert, "Disclosure-limited data dissemination," *Journal of The American Statistical Association*, pp. 10–28, 1986.

[24] D. Lambert, "Measures of disclosure risk and harm", *Journal of Official Statistics*, vol. 9, pp. 313–331, 1993.

[25] M. E. Nergiz, M. Atzori, and C. Clifton, "Hiding the presence of individuals from shared databases," in *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD)*, pp. 665–676, 2007.

[26] Anonymized Data: Generation, Models, *Usage Graham Cormode, Divesh Srivastava* AT&T Labs-Investigate, and Florham Park, NJ 07932 USA.

[27] J. Xu, W. Wang, J. Pei, X. Wang, B. Shi, and A.W.-C. Fu, "Utility- Based Anonymization Using Local Recoding,"
Proc. 12th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), pp. 785-790, 2006.

[28] K. LeFevre, D. DeWitt, and R. Ramakrishnan, "Workload-
Aware Anonymization," *Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), pp. 277-286, 2006.*

[30] C. Dwork, "Differential Privacy," *Proc. Int'l Colloquium Automata, Languages and Programming (ICALP), pp. 1-12, 2006.*

[31] C. Dwork, "Differential Privacy: A Survey of Results," *Proc. Fifth Int'l Conf. Theory and Applications of Models of Computation (TAMC), pp. 1-19, 2008.*

[32] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating Noise to Sensitivity in Private Data Analysis,"
Proc. Theory of Cryptography Conf. (TCC), pp. 265-284, 2006.

[33] R. Agrawal and R. Srikant, "Privacy preserving data mining,"
in *Proc. of ACM SIGMOD*, 2000.

[34] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "l-Diversity: Privacy Beyond k-Anonymity." in *Proc. of ICDE*, 2006.

[35] R. J. Bayardo and R. Agrawal, "Data Privacy through Optimal k- Anonymization," in *Proc. of ICDE*, 2005, pp. 217–228.

Authors:

1) Pargi Sridhar Reddy PG Scholar in Department of Computer Science and Engineering, TKR College of Engineering and Technology.

2) Ch. Ravi Associate Professor in Department of Computer Science and Engineering, TKR College of Engineering and Technology.

