www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 6 Issue 5 May 2017, Page No. 21258-21268
Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i4.12

Literature Survey of Privacy Preserving Data Publishing (PPDP) Techniques

Amita Sharma, N. Badal
Department of Computer Science & Engineering,
Kamla Nehru Institute of Technology,
Sultanpur, U.P, India.
safalta.amita@gmail.com, n_badal@hotmail.com

***Abstract*-Microdata-Information collected by different organizations is published for analysis to the analyst, decision makers, policy makers and researchers. Original data is not published due to some privacy issues. So, techniques are needed to preserve privacy of data. This paper includes the comparative study of various techniques available to preserve the privacy of published data.**

Keywords – privacy preservation, privacy preservation data publication(PPDP)

I. INRODUCTION

Microdata-Information includes the data collected from public. This data is published by the collectors to third party for further operation. Anonymized form of data is published so that privacy of data remains preserve. Many authors proposed techniques to solve this problem. This paper summarizes most of the techniques of PPDP.

This paper is divided into four sections. First section, discuss about the background information that is needed for the study of PPDP techniques. Second section, defines various PPDP techniques in condense form. Third section, includes others contribution in this field. Last section, comparison of defined techniques is done on the basis of various parameters and future scope in this field.

II. BACKGROUND

In 1977, Dalenius in his paper [22] define privacy preservation as "access to the published data should not enable the adversary to learn anything extra about any target victim compared to no access to the database, even with the presence of any adversary's background knowledge obtained from other sources." Need of Privacy preservation is illustrated with the help of data sets -publicly available data (table 1) and hospital record (table 2).

Table 1: Publicly available data

NAME	AGE	GENDER	PINCODE
Arun	19	M	214121
Imli	21	F	214452
Shenu	33	M	216353

Table 2: Hospital record

AGE	GENDER	PINCODE	DISEASE	CURE
19	M	214121	Flu	Yes
20	F	244452	Flu	Yes
23	F	216355	AIDS	No

DOI: 10.18535/ijecs/v6i4.12

Row 1 of both table1 and table 2 contains same values for age, gender and Pincode columns. From this adversary, can easily get information that Arun is having disease flu which is cured.

Before Publishing data, it is collected by Data Holders, as shown in figure 1.

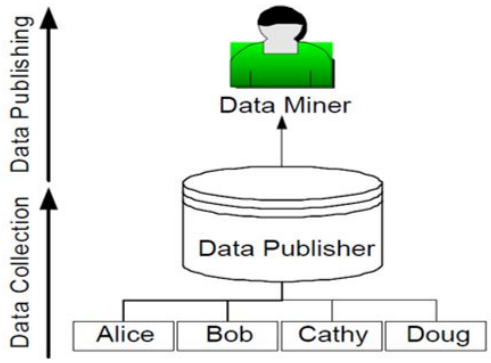

Figure 1: Data collection and Publication Phases

Data holders may be Trusted or untrusted. In case of trusted, data holder is trustworthy. In case of Untrusted, data holder may try to find the sensitive information from the collected data.

Categories of attributes

Attributes of data records hold by data holder can be categorized as identifier, quasi identifier, sensitive attributes, and non-sensitive attributes. Identifier are the set of attribute values that are publicly available and explicitly identify record owner e.g. Name. Quasi identifier(qid) are the set of attribute values that could potentially identify the record of owner. Sensitive attributes(sa) are the set of attributes that contains person specific sensitive information e.g. salary, disease. Non-sensitive attributes are all the remaining attributes of record.

Attack models

Attacks can be done on published data. So, Attack models is categorized into linkage models and probabilistic attack. Linkage models occurs when adversary is able to link published data with other data and get some information from the given data. It is done in three ways. These ways are Record linkage, Attribute linkage and Table linkage. Probabilistic attack occurs when adversary know some background information about the victim and he is able to get new information from the table about victim.

Types of privacy preservation

Privacy preservation is divided into following types: -

- Privacy preservation data mining(PPDM):PPDM uses tools and techniques of data mining. It modifies data to mask the sensitive information.in this Data recipient could be an adversary. It Directly hides sensitive data and Fails to preserve the truthfulness at record level.
- Privacy preservation data publishing(PPDP):In PPDP, sensitive data is not hide but hide the identity of an individual by anonymize the data.
- Privacy preserving distributed data mining(PPDDM): Data mining task is done by different placed by different parties and then combined the data and then published.
- Privacy preserving social network data publication [7]: Social network such as Facebook, LinkedIn etc. are published while preserving data owner privacy.

III. EXISTING METHODS

Many privacy preservation data publishing techniques are proposed by many authors. These techniques are as follows: -

DOI: 10.18535/ijecs/v6i4.12

A. K-anonymity

K-anonymity is Proposed by Samarati and Sweeny in [10] as "if any record in the table has same qid, at least k-1 other records also have the value qid". It is a Record linkage model.in this Minimum equivalence group size on qid is at least k. Probability of linking a victim to a specific record through qid is at most 1/k. possible Attacks on published data are Homogeneity attack and Background attack. For example, table 3 is the original Data and table 4 is 3-anonymous Patient Data.

Table 3: Original Patient Data

JOB	GENDER	AGE	DISEASE
Professional	M	35	Hepatitis
Professional	M	36	Hepatitis
Professional	M	38	HIV
Artist	F	32	Flu
Artist	F	31	HIV
Artist	F	33	HIV
Artist	F	34	HIV

Table 4: 3-anonymous Patient Data

JOB	GENDER	AGE	DISEASE
Professional	M	35-40	Hepatitis
Professional	M	35-40	Hepatitis
Professional	M	35-40	HIV
Artist	F	30-35	Flu
Artist	F	30-35	HIV
Artist	F	30-35	HIV
Artist	F	30-35	HIV

B. L-diversity

L-diversity is proposed by Ashwin Machanavajjhala in [3] as "every qid group contains at least l well-represented sensitive attributes". It is an Attribute linkage model. Possible attacks on published data is Similarity attack. For example, table 5 is the original Patient Data and table 6 is 3- diverse Patient Data.

Table 5: Original Patient Data

PINCODE	AGE	SALARY	DISEASE
47601	22	20K	Gastric ulcer
47610	22	30K	Gastritis
47631	26	40K	Cancer
47900	42	50K	Gastris
47906	48	100K	Flu
47903	50	70K	Bronchitis

Table 6: 3-diverse Patient Data

PINCODE	AGE	SALARY	DISEASE
476**	2*	20K	Gastric ulcer
476**	2*	30K	Gastritis
476**	2*	40K	Cancer
4790*	≥40	50K	Gastris
4790*	≥40	100K	Flu
4790*	≥40	70K	Bronchitis

C. T-closeness

T-closeness is Proposed by Li et al [12] as "the distribution of a sensitive attribute in any group on qid to be close to the distribution of the attribute in the overall table."

D. (α-k) anonymity

(α-k) anonymity is Proposed by Wong et al. in [16] as "every qid in table T to be shared by at least k records and conf(qid->s) ≤α for any sensitive value s, where k and α are data holder specified thresholds". Where α is a real number between [0,1] and K is the positive integer. Published data is free from inference attack. For example, table 7 is the original Data and table 8 is (0.5,2) anonymous Data.

Table 7: Original Data

JOB	BIRTH	PINCODE	DISEASE
Professional	1942	4350	HIV
Professional	1951	4350	Flu
Professional	1960	5432	Flu
Artist	1945	5432	Fever
Artist	1955	4350	Flu
Artist	1961	4350	Fever

Table 8: (0.5,2) anonymous Data

JOB	BIRTH	PINCODE	DISEASE
*	*	4350	HIV
*	*	4350	Flu
*	*	5432	Flu
*	*	5432	Fever
*	*	4350	Flu
*	*	4350	Fever

E. (X, Y) anonymity

(X, Y) anonymity is Proposed by Wang and Fung in [9] as: "each value on X is linked to at least k distinct values on Y". Where X and Y are disjoint set of attributes. For example, table 9 is (3,1) anonymous Data.

Table 9: (3,1) anonymous Data

JOB	GENDER	AGE	PINCODE
Artist	M	28	2350
Lawyer	F	30	2450
Teacher	M	32	2560
Artist	M	28	2351
Lawyer	F	30	2451
Teacher	M	32	2561
Teacher	M	32	2760

F. ε-differential privacy

ε-differential privacy is proposed by Dwork [5] as: "A randomized function F ensures ε-differential privacy if for all data sets T1 and T2 differing on at most one record $\left|\ln\{P[F9(T1=s)]/p[F(T2=s)]\}\right| \leq \varepsilon$. For all s ε Range(F) and Range(F)= possible set of outputs of the random function F." It provides guarantee against adversaries with arbitrary. It is a Probabilistic model. Smaller the value of ε means unable to distinguish between two datasets. For example, table 10 is original data and table 11 is Differential Privacy Data.

Table 10: Original Data

NAME	HAS DIABETES
Ross	1
Monica	1
Joy	0
Phoebe	0
chandler	1

Table 11: Differential Privacy Data

NAME	HAS DIABETES
Ross	1
Monica	1
Joy	0
Phoebe	0
chandler	0

G. (X-Y) privacy

(X-Y) privacy is Proposed by Wang and Fung in [9]. It is the Combination of (X,Y) anonymity and (X,Y) linkability. This is Applied on multiple release scenario.

H. (k, e) anonymity

(k, e) anonymity is proposed by Zhang et al. in [14] it is used for numerical sensitive attributes. It Partition record into groups so that each group contains at least k different sensitive values with a range of at least e. for example, table 12 is original data and table 13 is (7,50) anonymous Data.

Table 12: Original Data

JOB	SEX	SALARY
Artist	F	30K
Artist	F	31K
Artist	F	30K
Artist	F	32K
Artist	F	35K
Artist	F	34K
Artist	F	33K
Artist	F	32K
Artist	F	35K
Artist	F	80K

Table 13: (7,50) anonymous Data

QID		SENSITIVE	COMMENT
JOB	SEX	SALARY	
Artist	F	30K	Sensitive
Artist	F	31K	Sensitive
Artist	F	30K	Sensitive
Artist	F	32K	Sensitive
Artist	F	35K	Sensitive
Artist	F	34K	Sensitive
Artist	F	33K	Sensitive
Artist	F	32K	Sensitive
Artist	F	35K	Sensitive
Artist	F	80K	Non-sensitive

I. (d,γ) privacy

(d,γ) privacy is Proposed by Rastogi et al. in [24] as "A reasonable tradeoff between privacy and utility can be achieved only when prior belief is small." Where d is the difference of the prior and posterior probabilities and γ is the record.

J. Distributional privacy

Distributional privacy is Proposed by Blum et al. in [2] it is used for non-interactive query model. A mechanism satisfied (α-β) distributional privacy if for any distribution over database elements D, with probability (1-β), two databases D1 and D2 consisting of n elements drawn without replacement from D. No one reveal extra information about the sample than what is inherent from sample. E.g. hospital record of a particular region having patient with disease X. Data of patients is released anonymously without revealing the name of hospital from where data come from.

K. (X-Y) linkability

(X-Y) linkablity is proposed by Fung in [9]. Here X and Y are the attributes of a table. In this case, When particular Y values probability is higher than 1/k, Y can be replaced by subset of some values of yi where y= {y1, y2, y3,……}.

L. Personalized privacy

Personalized privacy is Proposed by Xiao and Tao in [26]. This technique allows each owner to specify her own privacy level. E.g. if a patient A has disease HIV then he can change HIV with Infectious disease while another patient B

cannot.

M. C-t isolation

c-t isolation is Proposed by Chawle et al. [18] as "having access to the published anonymous data table should not enhance an adversary's power of isolating any record owner." It is Suitable for numerical data.

N. Generalization

Generalization is proposed by Pierangela Samarati in [13]. In this, qid values are replaced by less specified but semantically consistent values. It uses k anonymity. Steps include in the implementation of this technique are Identifier removal, Tuple partitioning and Transform qid in each bucket means generalize qid values. This technique has some Limitations. First, it uses k-anonymity which suffers from the curse of dimensionality. Second, the data analyst has to make the uniform distribution assumption that every value in each generalized set is equally possible. Third, Correlations between different attributes are lost. Possible Attacks on the published data are Background knowledge attack and Homogeneity attack. For example, table 14 is original data and table 15 is Generalized Data.

Table 14: Original Data

AGE	GENDER	PINCODE	DISEASE
22	M	47906	Dyspepsia
22	F	47906	Flu
33	F	47905	Flu
52	F	47905	Bronchitis
54	M	47302	Flu
60	M	47302	Dyspepsia
60	M	47304	Dyspepsia
64	F	47304	Gastritis

Table 15: Generalized Data

AGE	GENDER	PINCODE	DISEASE
[20-52]	*	4790*	Dyspepsia
[20-52]	*	4790*	Flu
[20-52]	*	4790*	Flu
[20-52]	*	4790*	Bronchitis
[54-64]	*	4730*	Flu
[54-64]	*	4730*	Dyspepsia
[54-64]	*	4730*	Dyspepsia
[54-64]	*	4730*	Gastritis

O. Bucketization

Bucketization is proposed by David J. Martin in [6]. It uses l-diversity. It separates qid and sa and then randomly permute sa values. Steps of implementing this technique are Remove identifiers from the data, Partition tuples into buckets and Separate the SAs from the QIs by randomly permuting the SA values in each bucket. This technique has some Limitations. First, it Does not prevent membership disclosure. Second, it Requires a clear separation between QIs and SAs. Third, Breaks the attribute correlations between the QIs and the SAs. Possible attacks on published data are Skewness attack and Similarity attack. For example, table 14 is original data and table 16 is Bucketized Data.

Table 16: Bucketized Data

AGE	GENDER	PINCODE	DISEASE
22	M	47906	Dyspepsia
22	F	47906	Flu
33	F	47905	Flu
52	F	47905	Bronchitis
54	M	47302	Flu
60	M	47302	Dyspepsia
60	M	47304	Dyspepsia
64	F	47304	Gastritis

P. Anatomy

Anatomy is proposed by X. Xiao in [25]. It uses

l-diversity. Steps of implementation are Partition tuples of microdata into several QI-groups, Create QI table and Create ST (SA table) which contains SA statistics for each QI group. It Removes problem of generalization about uniform distribution assumption. For example, table 17 is QI table and table 18 is SA Table of table 14 original data.

Table 17: QI Table

AGE	GENDER	PINCODE	GROUP-ID
22	M	47906	1
22	F	47906	1
33	F	47905	1
52	F	47905	1
54	M	47302	2
60	M	47302	2
60	M	47304	2
64	F	47304	2

Table 18: SA Table

GROUP-ID	DISEASE	COUNT
1	Flu	2
1	Dyspepsia	1
1	Bronchitis	1
2	Gastritis	1
2	Flu	1
2	Dyspepsia	2

Q. Slicing

Slicing is proposed by Tiancheng Li in [23]. It uses k-anonymity as well as l-diversity. It Partitions the data horizontally as well as vertically. It is Better than generalization and bucketization. Steps of implementation include Attribute Partitioning, Column generalization and Tuple partitioning. Limitation of this technique is that It cannot provide better data utility for an analyst. For example, table 19 is sliced data of table 14 original data.

Table 19: Sliced Data

(AGE, GENDER)	(PINCODE, DISEASE)
(22, M) (22, F) (33, F) (52, F)	(47905, Flu) (47906, Dyspepsia) (47905, Bronchitis) (47906, Flu)
(54, M) (60, M) (60, M) (64, F)	(47304, Gastritis) (47302, Flu) (47302, Dyspepsia) (47304, Dyspepsia)

R. Overlapping slicing

Overlapping slicing is proposed by Suman S. Giri and Nilav Mukhopadhyay in [21]. It is the Extended version of slicing. In this an attribute is duplicated in more than one columns. Steps of implementation include Attribute Partitioning, Column generalization and Tuple partitioning. Limitation of this technique is that Still some utility of data
is lost. For example, table 20 is overlapping sliced data of table 14 original data.

Table 20: Overlapping sliced Data

(AGE, GENDER, DISEASE)	(PINCODE, DISEASE)
(22, M, Flu) (22, F, Dyspepsia) (33, F, Bronchitis) (52, F, Flu)	(47905, Flu) (47906, Dyspepsia) (47905, Bronchitis) (47906, Flu)
(54, M, Gastritis) (60, M, Flu) (60, M, Dyspepsia) (64, F, Dyspepsia)	(47304, Gastritis) (47302, Flu) (47302, Dyspepsia) (47304, Dyspepsia)

S. (p+)-sensitive t-closeness

(p+)-sensitive t-closeness is proposed by Sowmyarani C N and Dr. G N Srinivasan in [20] as "The table T satisfies (p+)-sensitive, t-closeness property, if it satisfies t-closeness, and each Qi-group has at least p distinct sensitivity level of values for the sensitive attribute." its prime concern is to preserve

privacy. For example, table 21 is Original Data, Table 22 defines the sensitivity level for sensitive attribute and table 23 is (2+) sensitive, 0.2-closeness data.

Table 21: Original Data

Age	Pincode	Salary	Disease
47977	21	360000	Heart Attack
47901	57	430000	Heart Attack
47982	47	380000	Diabetes
47904	45	590000	Diabetes
47609	34	143000	Brain Tumour
47605	21	600000	Bladder Cancer
47654	23	360000	Brain Tumor
47609	30	650000	Brain Tumor
47604	10	230000	Flu
47602	45	230000	Gastritis
47678	50	160000	Neck Pain
47903	21	467000	Neck Pain

Table 22: sensitivity levels for attribute values

Sn	Disease attribute values	Sensitivity level
1	Brain Tumour, Bladder Cancer	Top level
2	Heart Attack	Middle level
3	Diabetes, Gastritis	Low level
4	Flu, Neck Pain	Poor level

Table 23: (2+) sensitive, 0.2-closeness

Pin code	Age	Salary	Disease
47*** 47*** 47*** 47***	>20 >20 >20 >20	6 LPA 4 LPA 2 LPA 3 LPA	Brain Tumor Heart Attack Gastritis Heart Attack
47*** 47*** 47*** 47***	>20 >20 >20 >20	6 LPA 5 LPA 1 LPA 3 LPA	Bladder Cancer Diabetes Brain Tumor
47*** 47*** 47*** 47***	>10 >10 >10 >10	1 LPA 2 LPA 4 LPA 3 LPA	Neck Pain Flu Neck Pain Brain Tumor

T. (k, l) diversity

(k, l) diversity is proposed by Qiyuan Gong, Junzhou Luo, Ming Yang, Weiwei Ni, Xiao-Bai Li in [15]. It uses k-anonymity and l-diversity.it is used for 1:M dataset. 1:M dataset are those in which record of an individual occur more than one time. Steps of implementation includes transformation, SA anonymization and QID anonymization and SA diversity. It provides better utility then other techniques. As it uses generalization in anonymization of qid values so, the limitations are same as that of generalization. For example, table 24 is the original 1:M data and table 25 is (3,3) diverse data.

Table 24: Original 1:M Data

Tuple id	PID	Age	Gender	Pin code	Disease
1(Bob)	1	18	M	12000	a1
2(Bob)	1	18	M	12000	a2
3(Bob)	1	18	M	12000	b2
4(David)	2	14	M	13000	b1
5(Tom)	3	21	F	21000	b2
6(Simon)	4	16	M	14000	c2
7(Daisy)	5	27	F	22000	a2
8(Daisy)	5	27	F	22000	b2
9(Alice)	6	28	F	21000	c1
10(Alice)	6	28	F	21000	c2

Table 25: (3, 3) diverse Data

Age	Gender	Pin code	Disease
[11,20] [11,20] [11,20]	M M M	[10001,15000] [10001,15000] [10001,15000]	<A, b2> <B> <C>
[21,30] [21,30] [21,30]	F F F	[20001,25000] [20001,25000] [20001,25000]	<B> <A, b2> <C>

IV. Others contribution

Some authors in different-different year also give their contributions in this field through surveys and suggestions. In 2009, Yan Zhoo et al., in [27], do survey on PPDP techniques and conclude that PPDP is at the stage of development. In the same year, Bee-chung Chen et al., in [4], also discuss the overview of some techniques. In 2010, Ninghui et al., in [11], proposes a measure for published data named closeness. In 2011, Ruichen et al., in [17], proposes top-down partitioning algorithm based on differential privacy for publishing set-value data. In 2012, Junquiang Liu in [8] define anonymization techniques like generalization and data utility matrix with some challenges and future direction for this field. In 2014, Yang Xu in [28] presented a survey of PPDP includes privacy preserving model for record linkage and anonymity operations. In the same year, ANK Zaman and Charlie Obimbo, in [1], suggest how PPDP is used for classify data. In the same year, another survey on PPDP is done by S. Gokila and Dr. P. Venkateswari in [19], in this they done a comparative study of some PPDP techniques.

V. CONCLUSION

Different techniques of PPDP is discussed in the above section. Comparison between different techniques on the basis of various parameters (taken as columns) and technologies (taken as rows), is listed in table 26.

Preserving Privacy and Utility of data is important, privacy for public and utility for analyst. In future, techniques may be developed that preserve maximum privacy with maximum utility of data. Second, some more matrices are needed that shows direct relation between percentage of utility and privacy of data directly.

Table 26: comparison between different privacy preservation data publishing(PPDP)

Parameter	**Author**	**Attack model**				PP	UP
PPDP techniques		RL	AL	TL	PA		
k-anonymity	Samarati and Sweeny	√				√	
l-diversity	Machanavajjhala	√	√			√	
t-closeness	Li		√		√	√	
(α, k) anonymity	Wong	√	√			√	
(X, Y) anonymity	Fung	√	√			√	√
E-differential privacy	Dwork				√	√	√
(X, Y) privacy	Wang and fung	√	√			√	√
(k, e) anonymity	Zhang		√			√	√
(d, γ) privacy	Rastogi			√	√	√	
Distributional privacy	Blum		√	√		√	√
(X-Y) linkability	Fung	√	√			√	√
Personalized privacy	Xiao and Tao		√			√	
c-t isolation	Chawle	√			√	√	
Generalization	Pierangela Samarati	√				√	
Bucketization	David J. Martin	√	√			√	√
Anatomy	X. Xiao	√	√			√	√
Slicing	Tiancheng Li	√	√			√	√

Overlapping slicing	Suman S. Giri and Nilav Mukhopadhyay	√	√			√	√
(p+)-sensitive t-closeness	Sowmyarani C N and Dr. G N Srinivasan					√	
(k, l) diversity	Qiyuan Gong, Junzhou Luo, Ming Yang, Weiwei Ni, Xiao-Bai Li	√				√	√

RL-record linkage, AL-attribute linkage, TL-Table linkage, PA-Probability attack, PP-Privacy preservation and UP-Utility Preservation

VI. REFRENCES

[1] A N K Zaman, Charlie Obimbo. Privacy Preserving Data Publishing: A Classification Perspective. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 5, No.9, 2014.page 129-134

[2] A. Blum, K. Ligett, and A. Roth. A learning theory approach to non-interactive database privacy. In Proc. of the 40th annual ACM Symposium on Theory of Computing (STOC), pages 609–618, Victoria, Canada, 2008.

[3] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. *l*-diversity: Privacy beyond k-anonymity. ACM Transactions on Knowledge Discovery from Data (TKDD), 1(1), March 2007.

[4] Bee-Chung Chen, Daniel Kifer, Kristen lefevre And Ashwin Machanavajjhala. Privacy-Preserving Data Publishing. Foundations and Trends in Databases.

[5] C. Dwork. Differential privacy. In Proc. of the 33rd International Colloquium on Automata, Languages and Programming (ICALP), pages 1–12, Venice, Italy, July 2006.

[6] 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), pages 126-135, 2007.

[7] Jemal H. Abawajy, Mohd Izuan Hafez Ninggal, and Tutut Herawan. Privacy Preserving Social Network Data Publication. IEEE communications surveys & tutorials, vol. 18, no. 3, third quarter 2016.pages 1974-1997

[8] Junqiang Liu. Privacy Preserving Data Publishing: Current Status and New Directions. Information Technology Journal 11 (1): 1-8, 2012.

[9] K. Wang and B. C. M. Fung. Anonymizing sequential releases. In Proc. of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD), pages 414–423, Philadelphia, PA, August 2006.

[10] L. Sweeney. k-Anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-based Systems, 10(5):557–570, 2002.

[11] Li, Ninghui, Tiancheng Li, and Suresh Venkatasubramanian. "Closeness: A new privacy measure for data publishing." IEEE

Transactions on Knowledge and Data Engineering 22.7 (2010): 943-956.

[12] N. Li, T. Li, and S. Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and "-diversity. In Proc. of the 21st IEEE International Conference on Data Engineering (ICDE), Istanbul, Turkey, April 2007.

[13] P. Samarati, "Protecting Respondent's Privacy in Microdata Release," IEEE Trans. Knowledge and Data Eng., vol. 13, no. 6, pages 1010-1027, Nov./Dec. 2001.

[14] Q. Zhang, N. Koudas, D. Srivastava, and T. Yu. Aggregate query answering on anonymized tables. In Proc. of the 23rd IEEE International Conference on Data Engineering (ICDE), April 2007.

[15] Quyuan Gong, Junzhou Luo, Ming Yang, Weiwei Ni and Xiao Baili. Anonymizing 1:M microdata with high utility. Elsevier, Knowledge-Based-Systems, pages 1-12, October 23,2016.

[16] R. C. W. Wong, J. Li., A. W. C. Fu, and K. Wang. (α, k)-anonymity: An enhanced k-anonymity model for privacy preserving data publishing. In Proc. of the 12th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), pages 754–759, Philadelphia, PA, 2006.

[17] Rui Chen, Noman Mohammed, Benjamin C. M. Fung, Bipin C. Desai and Li Xiong. Publishing Set-Valued Data via Differential Privacy. Proceedings of the VLDB Endowment 4.11 (2011): 1087-1098.

[18] S. Chawla, C. Dwork, F. McSherry, A. Smith, and H. Wee. Toward privacy in public databases. In Proc. of Theory of Cryptography Conference (TCC), pages 363–385, Cambridge, MA, February 2005.

[19] S.Gokila, Dr.P.Venkateswari. A survey on privacy preserving Data publishing. International Journal on Cybernetics & Informatics (IJCI) Vol. 3, No. 1, February 2014

[20] Sownyarani CN and Dr. GN Srinivasan. A Robust Privacy Preserving Model for data Publishing. ICCCI-2015,2015.

[21] Suman S. Giri1and Mr.Nilav Mukhopadhyay. Overlapping Slicing with New Privacy Model. In International Journal of Scientific and Research Publications, Volume 4, Issue 6, June 2014. Pages 1-5.

[22] T. Dalenius. Towards a methodology for statistical disclosure control, Statistik Tidskrift. 15:429-444, 1997.

[23] Tiancheng Li, Ninghui Li, Senior Member, IEEE, Jian Zhang, Member, IEEE, and Ian Molloy. Slicing: A New Approach for Privacy Preserving Data Publishing. IEEE transactions on knowledge and data engineering, vol. 24, no. 3, march 2012., pages 561-574.

[24] V. Rastogi, D. Suciu, and S. Hong. The boundary between privacy and utility in data publishing. In Proc. of the 33rd International Conference on Very Large Data Bases (VLDB), pages 531–542, Vienna, Austria, September 2007.

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

[26] X. Xiao and Y. Tao. Personalized privacy preservation. In Proc. Of

ACM International Conference on Management of Data (SIGMOD), Chicago, IL, 2006.

[27] Yan Zhao1 Ming Du2 Jiajin Le1 Yongcheng Luo1. A Survey on Privacy Preserving Approaches in Data Publishing. 2009 First International Workshop on Database Technology and Applications.

[28] Yang Xu, Tinghuai Ma, Meili Tang and Wei Tian. A Survey of Privacy Preserving Data Publishing using Generalization and Suppression. Applied Mathematics & Information Sciences.2014. Pages 1103-1116

AUTHORS

Amita Sharma is a Research scholar in the Department of Computer Science & Engineering at Kamla Nehru Institute of Technology (KNIT), Sultanpur (U.P), INDIA. She received the B. Tech. in computer science and engineering in 2014 from Moradabad Institute of Technology, Moradabad.

N. Badal is a Sr. Lecturer in the Department of Computer Science & Engineering at Kamla Nehru Institute of Technology (KNIT), Sultanpur (U.P), INDIA. He received B.E. (1997) from Bundelkhand Institute of Technology (BIET), Jhansi in Computer Science & Engineering, M.E. (2001) in Communication, Control and Networking from Madhav Institute of Technology and Science (MITS), Gwalior and PhD (2009) in Computer Science & Engineering from Motilal Nehru National Institute of Technology (MNNIT), Allahabad. He is Chartered Engineer (CE) from Institution of Engineers (IE), India. He is a Life Member of IE, IETE, ISTE and CSI-India. He has published about 30 papers in International/National Journals, conferences
and seminars. His research interests are Distributed System, Parallel Processing, GIS, Data Warehouse & Data mining, Software engineering and Networking.