© 2018 IJSRST | Volume 4 | Issue 2 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

An Analytical Study on Privacy Preserving Data Publishing using Generalization and Suppression

Meeta B. Fadnavis
Lecturer, Department of Computer Management Dharampeth Polytechnic, Nagpur, Maharashtra, India

ABSTRACT

These days, information sharing as an imperative part shows up in our vision, realizing a mass of dialogs about strategies and systems of privacy preserving data publishing which are viewed as solid assurance to keep away from information exposure and ensure people's privacy. Late work concentrates on proposing diverse anonymity calculations for shifting data publishing situations to fulfill privacy prerequisites, and keep data utility in the meantime. K-anonymity has been proposed for privacy preserving data publishing, which can avoid linkage attacks by the methods for anonymity operation, for example, generalization and suppression. Various anonymity calculations have been used for accomplishing k-anonymity. This paper gives an outline of the advancement of privacy preserving data publishing, which is confined to the extent of anonymity calculations utilizing generalization and suppression. The privacy preserving models for attack is presented at first. A diagram of a few anonymity operations take after behind. The most vital part is the scope of anonymity calculations and information metric which is fundamental element of calculations. The conclusion and point of view are proposed at long last.

Keywords: Data Publishing, Privacy Preserving, Anonymity Algorithms, Information Metric, Generalization, Suppression

I. INTRODUCTION

Because of the quick development of information, the requests for data accumulation and sharing increment strongly. An incredible amount of data is utilized for investigation, measurements and calculation to discover general example or guideline which is helpful to social advancement and human advance. In the interim, dangers show up when huge data accessible for people in general. For instance, individuals can burrow privacy information by getting together sheltered appearing data, thusly, there is an awesome plausibility uncovering people privacy. As per the review, around 87 % of the number of inhabitants in the United States can be interestingly distinguished by given dataset distributed for people in general. To dodge this circumstance deteriorating, measures are taken by security division of numerous nations, for instance, declaring privacy control (e.g. privacy direction as a major aspect of Health Insurance Portability and Accountability Act in the USA [1]). The necessity for data distributer is that data to be distributed must fit for the predefined conditions. Distinguishing credit should be precluded from distributed dataset to ensure that people privacy can't be derived from dataset specifically. Expelling identifier characteristic is recently the readiness work of data handling, a few purification operations should be done further. In any case, after data preparing, it might diminish data utility significantly, while, data privacy did not get completely safeguarded.

In face of the testing risk, some explores have been proposed as a cure of this awkward circumstance, which focus at finishing the adjust of data utility and information privacy when publishing dataset. The progressing examination is called Privacy Preserving Data Publishing (PPDP). In the previous couple of years, specialists have taken up the test and undertaken a ton of looks into. Numerous possible methodologies are proposed for various privacy preserving situation, which fathom the issues in PPDP adequately. New strategies and hypothesis turned out persistently in specialists' push to finish privacy preserving.

A. Privacy Preserving Data Publishing

By and large, the procedure of Privacy Preserving Data Publishing has two stages; data gathering and data distribute stage. It alludes to three kinds of parts in the process who are data proprietor, data distributer and data

beneficiary. The relationship of two stages and three parts required in PPDP is appeared in figure 1. In the data accumulation stage, data distributer gathers dataset from data proprietor. At that point, in the data publishing stage, data distributer sends the prepared dataset to data beneficiary. It is important to say that crude dataset from data proprietor can't be straightforwardly sent to data beneficiary. The dataset ought to be prepared by data distributer before being sent to data beneficiary.

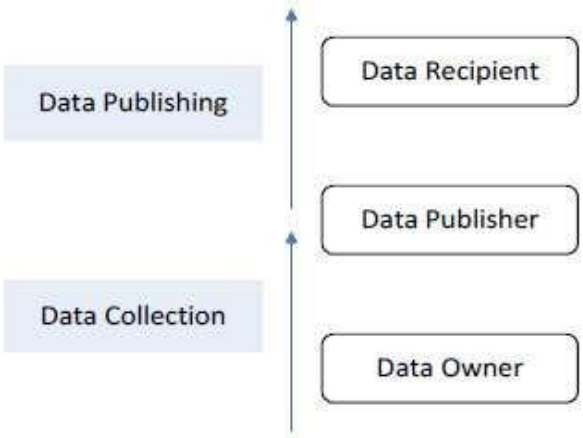

Figure 1. The relationship of phases and roles in PPDP

In [2], data distributer can be partitioned into two classes. In the untrusted demonstrate, data distributer is tricky who will probably pick up privacy from dataset. In the put stock in model, data distributer is solid and any data in their grasp is protected and with no risk. Inferable from the distinction of data publishing situations influenced by differing presumptions and necessities to data distributer, data beneficiaries purposes and different elements, it gives four situations for further point by point exchange that possibly show up in genuine privacy preserving data publishing in [3].

The principal situation is the non-master data distributer. In this situation, data distributer does not need particular knowledge about research fields. What they have to do is make data be distributed fulfilling the prerequisites of data utility and information privacy. The second one is the data beneficiary could be an attacker. This situation is all the more ordinarily acknowledged and many proposed arrangements make it as the imperative theory. The third one is the distribute data is not the data mining result. It demonstrates that dataset given by data distributer in this situation is not simply for data mining. That is to state, distributed dataset is not data mining result. The last one is honesty at record level. Data distributer ought to ensure the legitimacy of data to be distributed whatever preparing techniques will be utilized. Along these lines, randomization and bother can't meet the necessities in this situation.

B. K-Anonymity

When alluding to data anonymization, the most widely recognized data is two-dimensional table in social database. For privacy preserving, the characteristics of table are isolated into four classes which are identifier, semi identifiers, non-semi qualities and touchy property. Identifier can particularly speak to a person. Clearly, it ought to be evacuated before data handling. Semi identifiers are a particular succession of characteristics in the table that noxious attackers can take preferred standpoint of these qualities linking discharged dataset with other dataset that has been as of now procured, then breaking privacy, in the end increasing touchy information. Data cleansing worked by data distributer fundamentally focuses on semi identifiers. Because of vulnerability of the quantity of semi identifiers, each approach of PPDP accepts the semi identifiers arrangement ahead of time. Just along these lines can the accompanying handling do? Non-semi characteristics have less impact on data handling. Hence, now and then, these traits does not turn up in the advance of data preparing which enormously diminish memory use and enhance the execution of the proposed calculation. Touchy property contains delicate information, for example, sickness, pay. From table 1(2), this is a two-dimensional table to be distributed. As per above presentation, we can get the conclusion that ID is identifier. In the event that table 1(1) is a known table which attacker will use as background knowledge, then we know Birthday, Sex and ZipCode are semi identifiers, Work is non-semi trait and Disease is delicate characteristic.

From the case above, we know why data handling steps chiefly work on semi identifiers. Just along these lines would we be able to diminish the connection of dataset to be distributed and other dataset. In PPDP, the advance of data preparing is called data anonymization. K-anonymity is one of anonymization methodologies proposed by Samarati and Sweeney[4] that each record in dataset can't be recognized with at any rate another (k-1) records under the projection of semi identifiers of dataset after a progression of anonymity operations (e.g. supplant particular incentive with general esteem). K-

anonymity guarantees that the likelihood of particularly speaking to a person in discharged dataset won't extraordinary than 1/k. For instance in table 1, we find out about Miss Yoga has diabetes by linking registration data table with patient data table by Birthday, Sex and ZipCode characteristics notwithstanding expelling identifier. Imagine a scenario where it can't extraordinarily decide a record. Accordingly attacker has no capacity to recognize delicate information with full certainty. How to make quiet table in Table 1 meet 2-anonymity? One of viable ways is that supplanting data with year for Birthday characteristic and utilizing * supplant the last two character of ZipCode quality.

Table 1. Illustrate anonymization and k-anonymity

Name	Birthday	Sex	ZipCode
Myron	1990/10/01	Male	210044
Yoga	1980/05/22	Female	210022
James	1782/06/23	Male	210001
Sophie	1992/03/12	Female	210012

(1) Census Data

ID	Work	Birthday	Sex	ZipCode	Diease
231001	Student	1990/10/01	Male	210044	Cardipathy
231002	Clerk	1980/05/22	Female	210022	Diabetes
231003	Official	1990/08/12	Male	210021	Flu
231004	HR	1980/02/25	Female	210012	Caner

(2) Patient Data

K-anonymity has been widely examined lately [5, 6, 7, and 8]. After 2-anonymity, it can't deduce that Miss Yoga has diabetes, or perhaps she has tumor. Since in patient data table, there are two records that can be linked to one record in enumeration data table about Miss Yoga. We can see that k-anonymity effectively affects this situation.

II. PRIVACY PRESERVING MODEL FOR ATTACKS

The thorough meaning of privacy assurance by Dalenius [9] is that tending to the distributed dataset ought not to build any plausibility of enemy to increase additional information about people, even with the nearness of background knowledge. In any case, it is difficult to quantize the extent of background knowledge. In this way, a straightforward theory taken by numerous PPDP written works is that foe has restricted background knowledge. As indicated by enemies' attack standard, attack model can be characterized into two classes, which are linkage attack and probabilistic attack.

A. Privacy Model for Attacks

The linkage attack is that foe takes touchy information by the methods for linking with discharged dataset. It has three sorts of linkage, record linkage, trait linkage and table linkage. Semi identifiers are known by enemy in advance is the regular normal for linkage attack. Moreover, foe likewise gets a handle on the fundamental information of people and needs to know their delicate information under the situations of record linkage and quality linkage. While, table linkage attack puts more stresses on the point that whether known person's information introduces in discharged dataset. The privacy model of record linkage will be intricately depicted in the following segment, which is the critical piece of this paper.

For the attack of trait linkage, the enemy could derive delicate information from the discharged dataset in view of the dispersion of touchy incentive in the gathering that the individual has a place with. A fruitful induction is conceivable working on the distributed dataset that fulfills the capabilities of k-anonymity. The regular successful answer for the ascribe linkage attack is to reduce the connection of semi identifiers and delicate traits of unique dataset. Surely, others shows additionally sprout as of late to capture this kind of attack, like ℓ-diversity[10] and recursive (c, ℓ)-diversity[11], (X, Y)- Anonymity[12], (a, k)-Anonymity[13], (k, e)- Anonymity[14], t-closeness by Li et al.[15], customized privacy by Xiao and Tao[16] et cetera. Table linkage is unique in relation to both record linkage and characteristic linkage. In the table linkage attack, the nearness or nonappearance of individual record in discharged table has as of now uncovered the delicate information of the particular person. Nergiz et al. proposed the hypothesis of d - nearness to counteract table linkage and further bound the likelihood surmising event of individual record inside a given range [17].

The probabilistic attack can be portrayed in the situations that enemy won't instantly scratch delicate information from discharged dataset, while, the discharged dataset can help out for foe through expanding his/her background knowledge to some degree. This kind of attack is called probabilistic attack

that it turns up a noticeable deviation for increasing touchy information subsequent to getting to the discharged dataset. Probabilistic attack is not like linkage attack which exactly knows singular information, then increase delicate information consolidated with existed background knowledge, yet it concentrates on changing enemy's probabilistic certainty of getting privacy information subsequent to securing distributed dataset. The privacy model to this attack needs to guarantee that the change of probabilistic certainty is moderately little in the wake of acquiring the distributed dataset. Some shrewd thoughts for probabilistic attack are (c, t)- isolation[18], e - differential privacy[19], (d, g)- privacy[20], distributional privacy[21] et cetera. Distinctive privacy preserving model has its novel components controlled by points of interest of the horrendous attack, so related calculations which have a place with a particular privacy model are altered and focused at settling specific attack circumstance.

B. Privacy Model for Record Linkage

For record linkage attack, we should find out about the meaning of identicalness class at first. At the point when the qualities under the projection of semi identifiers of dataset are same, the specific quantities of records frame a gathering. Many gatherings make up the dataset. Those gatherings are called identicalness class. In the first dataset, the measure of equality class changes drastically. In the event that attackers known record of discharged dataset coordinating a gathering with just a single record best case scenario circumstance, lamentably, the privacy information of individual identified with the just a single record will be leaked. For instance, the dataset in Table 2(1) should be discharged. In the event that publishing it without completing any anonymity operations and expecting that enemy has the background knowledge of Table 2(2). We can promptly find that Myron who is conceived in Nanjing, China on 1990 has normal cerebral pain by linking the two datasets in table 2 on Birthday, Sex and ZipCode. These three characteristics are called semi identifiers of this attack from the definition presented previously. K-anonymity is a strategy to explain record linkage attack which ensures that the measure of every identicalness class is more prominent or equivalent than the given esteem k by the methods for supplanting particular incentive with general esteem. The likelihood of interestingly deriving the delicate information of individual known by the enemy is under 1/k, accordingly, it can defend people's privacy to an expansive degree. Each semi identifier has a scientific categorization tree structure of which generalization degree increments from leaf to root hub. Experimentally, every straight out semi identifier has a foreordained scientific classification tree, while, the scientific classification tree of numerical semi identifier will be powerfully created in the execution of anonymity calculation, and, what's more, a particular estimation of numeric property will be supplanted by an all-around divided range in generalization. The scientific categorization tree structures of two semi identifiers are appeared in figure 2. For instance, in scientific classification tree structures of Job quality, the root hub ANY is more broad than hub Student. The parent hub Student is broader than its youngster hub Graduate.

There are numerous lovely techniques to take care of the issue of data anonymization, which obey capability of k-anonymity or its augmentation. The nitty gritty depiction will be presented in resulting section. With respect to k-anonymity, the majority of late works accept that there exists just a single semi identifier arrangement containing every conceivable trait. With the quantity of semi identifier expanding, not exclusively does it take more push to do one anonymity operation, additionally level of data mutilation increments separately. Thus, a few analysts propose a particular viewpoint taking multi semi identifier successions into record which is more adaptable than one semi identifier arrangement.

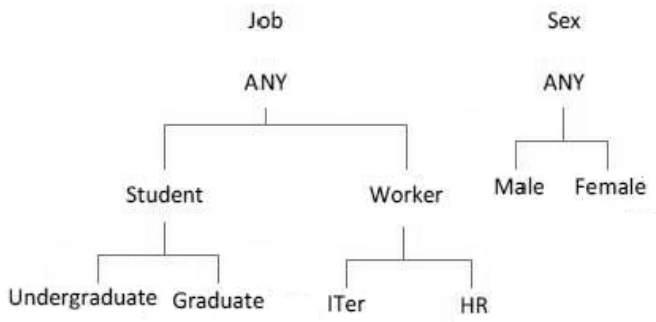

Figure 2. taxonomy tree structure of quasi-identifier

In any case, the way is picked; the assurance of properties in semi identifier needs many endeavours. No strategy or hypothesis can manage all issues in the particular research range. A short time later, expansions of k-anonymity are proposed. For example, (X, Y) - anonymity and Multi-Relational k-anonymity [22]. (X, Y)- anonymity has more strict limitations than k-

anonymity by annexing extra necessities, and it is for the situation that an individual is mapped to more than one record in discharged dataset, which implies that the unmistakable number of Y characteristic must more prominent or equivalent than the given k on the projection of X.

Table 2. Illustrate record linkage

Work	Birthday	Sex	ZipCode	Diease
Student	1990/10/01	Male	210044	Headache
Clerk	1980/05/22	Female	220022	Diabetes
Official	1990/08/12	Male	210021	Flu
HR	1980/02/25	Female	220012	Caner

(1) Patient Data

Name	Birthday	Sex	ZipCode
Myron	1990/10/01	Male	210044
Yoga	1980/05/22	Female	210022
James	1782/06/23	Male	210001
Sophie	1992/03/12	Female	210012

(2) Background Knowledge

Work	Birthday	Sex	ZipCode	Diease
Student	1990	Male	2100**	Headache
Clerk	1980	Female	2200**	Diabetes
Official	1990	Male	2100**	Flu
HR	1980	Female	2200**	Caner

(3) 2-anonymous patient data

Multi-Relational k-anonymity grows the limit of k-anonymity, which is for anonymizing numerous datasets rather than just a single dataset. Essentially, k-anonymity, (X, Y)- anonymity and Multi-Relational k-anonymity constitutes the hypothesis reason for privacy demonstrate for record linkage.

III. ANONYMITY OPERATIONS

A progression of anonymity operations works on unique dataset to make it satisfy the privacy prerequisite amid data anonymization. The every now and again utilized anonymity operations are generalization, suppression, anatomization, change and bother. Various calculations toward privacy preserving data publishing contrast in the decision of anonymity operations. Or, on the other hand, to place it in another way, the possibility of calculation depends on some particular anonymity operations.

A. Anonymity Operations

Generalization and suppression are the most widely recognized anonymity operations used to execute k-anonymity and its augmentation which are additionally delineated in the following session. Utilizing one sentence to clarify generalization is that supplanting particular estimation of semi identifiers with more broad esteem. Suppression is a definitive condition of generalization operation which utilizes extraordinary typical character to supplant its bona fide esteem (e.g. *, and, #), and makes the esteem good for nothing. Unlike generalization and suppression, anatomization and stage does not make any alteration of unique dataset, while diminish the relationship of semi identifiers and delicate characteristic. By and large, semi identifiers and delicate trait are distributed independently. Many examine make utilization of these two anonymity operations [23, 24, 25]. At the point when simply alluding to the motivation behind information measurement, bother operation has benefits of effortlessness and proficiency. The primary thought of annoyance is to substitute unique incentive for manufactured data, and, guarantees the factual normal for unique dataset. After annoyance operation, the dataset is totally not the introduction of unique dataset which is its remarkable attribute. Including clamour, swapping data and producing manufactured data are the three basic methods for annoyance [26, 27, 28, 29, 30].

B. Generalization and Suppression

Accomplishing k-anonymity by generalization and suppression will prompt not exact, but rather steady portrayal of unique dataset. Complete thought should be taken around three key angles alluded to PPDP, which are privacy necessity, data utility and algorithm complexity. There are approximately four sorts of generalization with contrast in degree and guideline which are full-area generalization, subtree generalization, cell generalization and multidimensional generalization. Coincidentally, determination is the invert anonymity operation of generalization.

1) Full-area generalization [31] is proposed in early research of PPDP, it has the littlest hunt space in four sorts of generalization, while it prompts vast data mutilation. The key of full-space generalization is that the estimation of semi identifier must be summed up to a similar level in given scientific categorization tree

structure. We will utilize the scientific categorization tree structure in figure 2 to clarify. Prior to any anonymity operation, all esteems remain at the base of scientific classification. On the off chance that hub Undergraduate is summed up to its parent hub Student, then hub Graduate must be summed up to hub Student, in the meantime, hubs ITer and HR should be summed up to hub Worker.

2) Subtree generalization [32, 33], its limit is littler than full-space generalization. At the point when a hub in scientific classification tree structure sums up to its parent hub, all tyke hubs of the parent hub should be summed up to the parent hub. For instance, in figure 2, if hub Undergraduate is summed up to its parent hub Student, it needs to sum up hub Graduate to its parent hub Student to meet the prerequisite of subtree generalization. Unhindered subtree generalization [34] is like the subtree generalization, aside from that kin of the summed up hub could stay unaltered. For instance, in figure 2, if hub Undergraduate is summed up to its parent hub Student, hub Graduate is pointless to sum up to its parent hub Student.

3) Cell generalization [35] is marginally unique in relation to generalization routes above. Cell generalization is for single record, while, full-area generalization is for all of records in the dataset. The inquiry space with this generalization is altogether bigger contrasted with other generalization, yet the data bending is moderately little. For instance, in figure 2, when hub Undergraduate sums up to its parent hub Student, it can keep up the record with Undergraduate esteem in the dataset. At the point when the mysterious dataset is utilized for grouping of data mining, it experiences data investigation issue. For instance, classifier may not know how to recognize Undergraduate and Student. Those issues are the basic qualities of nearby recoding plan.

4) Multidimensional generalization [34, 35] accentuates distinctive generalization for various blends of estimations of semi identifiers. For example, in fig 2, [Undergraduate, Female] can be summed up to [Student, ANY], while [Graduate, Male] sums up to [ANY, Male]. This plan has less data contortion contrasted with full space generalization. It sums up records by mix of semi identifiers with various esteem.

Much of the time of generalization plans; it blends suppression operations in its procedure of data anonymization. It is with no questions that there exists some hypothesis or methods to go on data anonymization just utilizing suppression operation. Like the classification of generalization, there are five kinds of suppression, characteristic suppression [34], record suppression, esteem suppression, cell suppression and multidimensional suppression. Trait suppression stifles the entire estimations of the quality. Record suppression implies smothering the records. Esteem suppression alludes to smothering the given an incentive in the dataset. While cell suppression contrasted with cell generalization, works on little extension and stifles a few records with the given an incentive in dataset.

IV. CONCLUSIONS

Information sharing is getting to be noticeably vital piece of people and associations, privacy preserving data publishing comes to get expanding considerations from everywhere throughout the world, which is viewed as a fundamental assurance for information sharing. Basically, the part of privacy preserving data publishing is to change the first dataset from one state to the next state to keep away from privacy revelation and withstand different attacks.

In this paper, to start with, we talk about privacy model of PPDP in points of interest, for the most part present privacy preserving model for record linkage and anonymity operations. Information metric with various objects are gathered which is a critical piece of anonymity calculations. In this way, more accentuation is put on the anonymization calculations with particular anonymity operations, precisely, generalization and suppression operation. This paper might be utilized for scientist to scratch the profile of anonymity calculations for PPDP by the methods for generalization and suppression.

Our further research appears underneath.

A. Hybrid k-anonymity calculation. Calculation usage of k-anonymity is basic and can adjust to various situations. In this way, it will be a compelling plan to blend k-anonymity with other anonymity strategies.

B. Background knowledge attack reproduction to make information safe. It is hard to precisely reenact the background knowledge of attackers. While diverse background knowledge will bring about privacy rupture in fluctuating degree. It will be a piece of research to discover the method for recreating background knowledge of attackers, so that give all-round insurance to privacy.

C. Information metric. It has given a general outline of information metrics in this paper. We can see that diverse metric fits for various situation of PPDP. Concentrate new information metric or enhancing existed metric will be a piece of further research.

D. Multi touchy traits anonymity limitation. Existing review concentrates on anonymization of a solitary delicate trait, which can't just move to take care of the multi touchy characteristic issue.

Along these lines, we have to concentrate successful anonymity calculations with multidimensional limitation. Furthermore, the challenges of actualizing customize anonymity proficiently and picking semi identifiers precisely all deserve of additionally thought and study.

V. REFERENCES

[1]. M. S. Wolf, C. L. Bennett, Local perspective of the impact of the HIPAA privacy rule on research, Cancer-Philadelphia Then Hoboken, 106, 474-479 (2006).

[2]. Johannes Gehrke, Models and methods for privacypreserving data publishing and analysis, In Proceedings of the 22nd International Conference on Data Engineering (ICDE), 105, (2006).

[3]. B. Fung, K. Wang, R. Chen, P. Yu, Privacy-preserving data publishing: A survey of recent developments, ACM Computing Surveys, 42, 1-53 (2010).

[4]. L. Sweeney, K-anonymity: A model for protecting privacy, International Journal on Uncertainty, Fuzziness, and Knowledge-based Systems, 10, 557-570 (2002).

[5]. Dewri R., k-anonymization in the presence of publisher preferences, Knowledge and Data Engineering, IEEE Transactions, 23, 1678-1690 (2011).

[6]. Nergiz M. E., Multirelational k-anonymity, Knowledge and Data Engineering, IEEE Transactions, 21, 1104-1117 (2009).

[7]. Jiuyong Li, Wong, R. C. W. Wai-chee Fu, A., Jian Pei, Transaction anonymization by local recoding in data with attribute hierarchical taxonomies, Knowledge and Data Engineering, IEEE Transactions, 20, 1181-1194 (2008).

[8]. Tamir Tassa, Arnon Mazza, k-Concealment: An Alternative Model of k-Type Anonymity, Transactions on Data Privacy, 189-222 (2013).

[9]. Dalenius T., Towards a methodology for statistical disclosure control, Statistik Tidskrift, 15, 429-444 (1977).

[10]. Ahmed Abdalaal, Mehmet Ercan Nergiz, Yucel Saygin, Privacy-preserving publishing of opinion polls, Computers & Security, 143-154 (2013).

[11]. Machanavajjhala A., Gethrke J., Kifer D., Venkitasubramaniam M., l-diversity: Privacy beyond kanonymity, In Proceedings of the 22nd IEEE International Conference on Data Engineering (ICDE), 24, (2006).

[12]. Ke Wang, Benjamin C. M. Fung, Anonymizing sequential releases, In Proceedings of the 12th ACM SIGKDD Conference, 414-423 (2006).

[13]. Raymond Chi-Wing Wong, Jiuyong Li, Ada Wai-Chee Fu, KeWang, (a, k)-anonymity: An enhanced k-anonymity model for privacy preserving data publishing, In Proceedings of the 12th ACM SIGKDD, 754-759 (2006).

[14]. Qing Zhang, Koudas N., Srivastava D., Ting Yu, Aggregate query answering on anonymized tables, In Proceedings of the 23rd IEEE International Conference on Data Engineering (ICDE), 116-125 (2007).

[15]. Ninghui Li, Tiancheng Li, Venkatasubramanian S., tcloseness: privacy beyond k-anonymity and l-diversity, In Proceedings of the 21st IEEE International Conference on Data Engineering (ICDE), 106-115 (2007).

[16]. Xiaokui Xiao, Yufei Tao, Personalized privacy preservation, In Proceedings of the ACM SIGMOD Conference, 229-240 (2006).

[17]. Mehmet Ercan Nergiz, Maurizio Atzori, Chris Clifton,Hiding the presence of individuals from shared databases, In Proceedings of ACM SIGMOD Conference, 665-676 (2007).

[18]. Chawla S., Dwork C.,Mcshrry F., Smith A.,Wee H., Toward privacy in public databases, In Proceedings of the Theory of Cryptography Conference (TCC), 363-385 (2005).
[19]. Dwork C., Differential privacy, In Proceedings of the 33rd International Colloquium on Automata, Languages and Programming (ICALP), 1-12 (2006).
[20]. Rastogi V., Suciu D., Hong S., The boundary between privacy and utility in data publishing, In Proceedings of the 33rd International Conference on Very Large Data Bases (VLDB), 531-542 (2007).
[21]. Blum A., Ligett K., Roth A., A learning theory approach to non-interactive database privacy, In Proceedings of the 40th Natural Sciences Publishing Cor.
[22]. Appl. Math. Inf. Sci. 8, No. 3, 1103-1116 (2014)/www.naturalspublishing.com/Journals.asp 1115 Annual ACM Symposium on Theory of Computing (STOC), 609-618 (2008).
[23]. Nergiz M. E., Clifton C., Nergiz A. E., Multirelational k-Anonymity, Knowledge and Data Engineering, IEEE Transactions on, 21, 1104-1117 (2009).
[24]. Xiangmin Ren, Research on privacy protection based on kanonymity, Biomedical Engineering and Computer Science (ICBECS), 1-5 (2010).
[25]. Vijayarani S., Tamilarasi A. Sampoorna M., Analysis of privacy preserving k-anonymity methods and techniques, Communication and Computational Intelligence (INCOCCI), 540-545 (2010).
[26]. Wenbing Yu, Multi-attribute generalization method in privacy preserving data publishing, eBusiness and Information System Security (EBISS), 1-4 (2010).
[27]. J. Domingo Ferrer, A survey of inference control methods for privacy preserving data mining, Advance in Database Systems, 34, 53-80 (2008).
[28]. N. Shlomo, T. De Waal, Protection of micro-data subject to edit constraints against statistical disclosure, Journal of Official Statistics, 24, 229-253 (2008).
[29]. I. Cano, V. Torra, Edit constraints on microaggregation and additive noise, In Proceedings of the International ECML/PKDD Conference on Privacy and Security Issues in Data Mining and Machine Learning (PSDML10), Springer, 1-14 (2010).
[30]. Jing Zhang, Xiujun Gong, Zhipeng Han, Siling Feng, An improved algorithm for k-anonymity, Contemporary Research on E-business Technology and Strategy Communications in Computer and Information Science, 352-360 (2012).
[31]. Xiaoling Zhu, Tinggui Chen, Research on privacy preserving based on k-anonymity, Computer, Informatics, Cybernetics and Applications Lecture Notes in Electrical Engineering, 107, 915-923 2012.
[32]. P. Samarati, Protecting respondents identities in microdata release, IEEE Trans. On Knowledge and Data Engineering, 13, (2001).
[33]. V. Iyengar, Transforming data to satisfy privacy constraints, In ACM SIGKDD, (2002).
[34]. Xuyun Zhang, Chang Liu, Surya Nepal, Jinjun Chen, An efficient quasi-identifier index based approach for privacy preservation over incremental data sets on cloud, Journal of Computer and System Sciences, 542-555 (2013).
[35]. Lefevre K., Dewitt D. J., Ramakrishnan R., Incognito: efficient full-domain k-annymity, In Proceedings of ACM SIGMOD, 49-60 (2005).
[36]. Xu J.,WangW., Pei J.,Wang X., Shi B., Fu A.W. C., Utilitybased anonymization using local recoding, In Proceedings of the 12th ACM SIGKDD Conference,(2006).