Protecting the Publishing Identity in Multiple Tuples

Youdong Tao, Yunhai Tong, Shaohua Tan, Shiwei Tang, and Dongqing Yang

Key Laboratory of Machine Perception (Peking University), Ministry of Education, 100871 Beijing, China
{taoyd,yhtong,shtan,tsw,dqyang}@pku.edu.cn

Abstract. Current privacy preserving methods in data publishing always remove the individually identifying attribute first and then generalize the quasi-identifier attributes. They cannot take the individually identifying attribute into account. In fact, tuples will become vulnerable in the situation of multiple tuples per individual. In this paper, we analyze the individually identifying attribute in the privacy preserving data publishing and propose the concept of identity-reserved anonymity. We develop two approaches to meet identity-reserved anonymity requirement. The algorithms are evaluated in an experimental scenario, demonstrating practical applicability of the approaches.

Keywords: Privacy preservation; Data publishing; Anonymity; Identity

1 Introduction

In recent privacy preserving data publishing research, k-anonymity principle [8,9,10] is of importance. It first removes the individually identifying attribute, then generalizes the quasi-identifier attributes and divides the tuples into different groups. It guarantees that each group has at least k tuples and the tuples in one group share the same quasi-identifier attribute values after generalization. Other enhanced principles, such as l-diversity [7], (α,k)-anonymity [12], extend this basic idea. All these methods have a default precondition that each individual has at most one tuple in the data set. In some real circumstances, that precondition doesn't meet.

For example, in a patient dataset published by a hospital (Table 1), some persons may appear more than one time for different diseases. In Table 1, Mike appears twice for two diseases: hypertension and hyperlipemia. If a number of people get both diseases at the same time, we may draw a conclusion that these two diseases are related. If the individually identifying attribute is removed, we can't make such a conclusion.

Current methods first remove the Name attribute and then generalize the quasi-identifier attributes. If we set k=2 in k-anonymity (or l=2 in l-diversity or α=0.5, k=2 in (α,k)-anonymity), we will get the result table as Table 2. From Table 1, we notice the first 2 rows will be grouped together since they share

the same quasi-identifier attribute values and have different diseases. According the k-anonymity assumption, if an adversary notices that Mike belongs to the Group 1, the probability that the adversary reveals the Mike's disease should be 50%. In fact, whatever disease the adversary judges, it is true and probability of privacy breach is 100%. This defect appears because they ignore the condition that one person may appear several times in a dataset.

Table 1. A patient table in which someone appears more than once

No	Name	Sex	Postcode	Disease
1	Mike	M	10085	hypertension
2	Mike	M	10085	hyperlipemia
3	Emily	F	10075	diabetes
4	Tim	M	10075	heart
5	Jane	F	10086	cancer
6	Ella	F	10087	flu

Table 2. Published table after common generalization

Group id	Sex	Postcode	Disease
1	M	10085	hypertension
1	M	10085	hyperlipemia
2	*	10075	diabetes
2	*	10075	heart
3	F	1008*	cancer
3	F	1008*	flu

This paper analyzes this situation and proposes an identity-reserved anonymity method. It modifies the current anonymity principles and reserves more information. The contributions of this paper are:

- We propose 3 identity-reserved anonymity principles. These principles reserve more information inside the dataset while protecting the individual privacy. The current anonymity principles don't take the multiple tuples per individual into account.
- We implement two algorithms to achieve identity-reserved anonymity principles. Global recoding algorithm extends the Incognito [4] to solve this problem. For less distortion, we adopt the domain generalization with tuple suppression. We also propose a local recoding algorithm and achieve less distortion. The algorithms are evaluated in an experimental scenario, demonstrating practical applicability of the approaches.

The remainder of this paper is organized as follows. In section 2, we review the related work before. In section 3, we propose the identity-reserved anonymity principles. In section 4, we discuss algorithms to implement these principles. We present experimental result in section 5 and conclude in section 6.

2 Related Work

In recent years, privacy preserving data publishing has gotten widely researched. Samarati and Sweeney proposed a principle called k-anonymity [8,9,10]. That requires each tuple in the table should be indistinguishable from at least (k-1) other tuples with respect to every set of quasi-identifier attributes. Beyond the k-anonymity, Machanavajjhala et al. proposed *l*-diversity principle [7]. That requires each quasi-identifier group should have at least l "well-represented" sensitive values. That principle extends the k-anonymity and diversifies the sensitive attribute values. They provided multiple interpretations on "well-represented". A simple interpretation on "well-represented" is that each quasi-identifier group has l distinct sensitive attribute values. Wong et al. proposed (α,k)-anonymity [12]. That requires each sensitive value in a quasi-identifier group should appear no more than a fixed frequency α besides k-anonymity. Li et al. proposed t-closeness principle which requires that distribution of sensitive attribute in groups should be close [6]. All these methods first remove the individually identifying attribute and generalize quasi-identifier attributes. Xiao and Tao proposed a personalized anonymity [13]. When they analyzed the probability of privacy breach, they distinguished two cases. One is the primary key scenario which each individual appears at most once. The other is the non-primary key scenario which each individual may appear an arbitrary number of times.

Generalization and suppression are the main approaches to achieve anonymity principles. Generalization is to replace a detailed value by a general value. Suppression is to delete some tuples. In generalization processing, suppression may be adopted. Suppression helps to decrease the generalization degree.

There are two main models in the algorithm of anonymity. One is global recoding [3,4,8,9], and the other is local recoding [1,9]. In global recoding, all values of an attribute should be generalized to the same domain level in hierarchy. But it always suffers from over-generalization and loses much information. In local recoding, values of an attribute may be generalized to different levels in hierarchy. For example, Table 2 is a result table of local recoding because some values of SEX attribute keep unchanged while some are generalized to the unknown value (*). If a recoding model divides an attribute into a set of non-overlapping intervals, it is called a single-dimensional recoding [3,4,8,9]. On the other hand, multidimensional recoding [5] divides the domain into a set of non-overlapping multidimensional regions. Besides generalization, Xiao and Tao proposed an "anatomy" method to meet the privacy requirement [14]. They published a quasi-identifier table (QIT) and a sensitive table (ST). These two tables share the same group-id attribute. In fact, it is a lossy join of database table.

3 Conceptions and Ideas

3.1 Identity Processing

The attributes of original table are classified to 3 types: (1) Individually identifying attribute (ID), that explicitly indicates an individual, such as name, SSN and mobile number. (2)Quasi-identifier attribute (QI), that can be exploited for linking and of k-anonymity as characterizing the degree of data protection with respect to inference by linking, such as sex, age and zip code. (3) Sensitive attributes (ST), that describe the privacy information of an individual, such as disease or income.

Removal of the ID attributes is the first step in common methods. But that processing loses individual information and may lead to the privacy breach, as shown before. We propose to recode and reserve the ID attribute for publishing. Recoding the ID attribute values is simply to replace it by a randomized number or string.

Reserving the ID attribute dramatically improves the utility of data set. For example in Table 2, reserving the ID attribute could help the research of complicating diseases which often appear together such as hypertension and hyperlipemia.

Individually identifying attributes may be specified by the publisher. We recode one of them and discard other individually identifying attributes since they are of redundancy.

3.2 Identity-reserved Anonymity

We reserve the individually identifying attribute and propose identity-reserved anonymity. In common k-anonymity, there are at least k tuples in every set of quasi-identifier attribute values. Similarly, in identity-reserved k-anonymity, there are at least k individuals in every set of quasi-identifier attribute values.

Definition 1 (Identity-reserved k-anonymity requirement). *Every release version of data must be such that every combination of values of quasi-identifiers can be indistinctly matched to at least k different individuals.*

This definition is the same as the notion in [8]. But in [8], it takes it regarded that each tuple links with a distinct individual. So it removes the explicit ID attribute at first. With ID attribute recoded, we define a data requirement based on individuals.

In the previous papers, published table's format is $T(QI, ST)$. QI is the combination of quasi-identifier attributes and ST is the sensitive attribute. In this paper, published table's format is $T(ID, QI, ST)$. ID is the recoded identifier, QI is the combination of quasi-identifier attributes and ST is the sensitive attribute. Let A= $\{a_1, a_2, \ldots, a_b\}$ be the individual set of T.ID and S=$\{s_1, s_2, \ldots, s_t\}$ be the distinct sensitive values set of T.ST. For each a_i, $S(a_i)$ is the sensitive attribute value set associated with the individual a_i. For each s_j, $A(s_j)$ is the individual set associated with the sensitive value s_j. QI consists of one or several quasi-identifier

attributes. The tuples shared the same combination of values of quasi-identifiers after generalization form a QI group. In a QI group Q, let $m = |\bigcup_{a_i \in Q.ID} S_{a_i}|$ and $n = |\bigcup_{s_j \in Q.ST} A_{s_j}|$.

Definition 2 (Identity-reserved k-anonymity). *Let T(ID, QI, ST) be a published table and QI be a quasi-identifier associated with it. ID is the recoded identifier, ST is the sensitive attribute. T is said to satisfy identity-reserved k-anonymity with regard to QI if each sequence of values in T.QI appears at least with k distinct occurrences in T.ID. That is in any QI group Q,* $n = |\bigcup_{s_j \in Q.ST} A_{s_j}| \geq k$.

For protecting the sensitive attributes, *l*-diversity is proposed. A naive interpretation of *l*-diversity requires that each QI group should have l different sensitive values. *l*-diversity principle doesn't take the situation into account that an individual may correspond to several tuples in the published table.

Definition 3 (Identity-reserved (*k,l*)-diversity). *Let T(ID,QI,ST) be a published table and QI be a quasi-identifier associated with it. ID is the recoded identifier, ST is the sensitive attribute. T is said to satisfy identity-reserved (k,l)-diversity if any QI group Q satisfies* $m = |\bigcup_{a_i \in Q.ID} S_{a_i}| \geq l$ *and* $n = |\bigcup_{s_j \in Q.ST} A_{s_j}| \geq k$.

(α,k)-anonymity takes the sensitive attribute value frequency into account. It requires that in each QI group, every sensitive value frequency should be no more than α and the size of each QI group should be no less than k. In our context, we propose identity-reserved (α,β)-anonymity. (α,β)-anonymity requires the frequency of sensitive and individually identifying attribute value in a QI group. Since it requires the frequency of individually identifying attribute, parameter k is abandoned.

Definition 4 (Identity-reserved (α,β)-anonymity). *Let T(ID,QI,ST) be a published table and QI be a quasi-identifier associated with it. ID is the recoded identifier, ST is the sensitive attribute. T is said to satisfy identity-reserved (α,β)-anonymity if in any QI group , each individual frequency is no more than α, and each sensitive value frequency is no more than β,* $0 < \alpha, \beta < 1$.

3.3 Privacy Breach Probability

In this section, we analyze the probability of privacy breach. In [13] the situation was discussed that an adversary has an external database for linking without any other background knowledge. Now we only discuss the situation that the adversary confirms someone (called "T") in the published table and knows T's QI attribute values. So the adversary knows the group that T belongs to (called group "G").

In group G, let individual set be $\{a_1, \ldots, a_n\}$. a_i appears c_i times in G, i=1,...,n. Assume $c_1 \geq c_2 \geq \ldots \geq c_n$. In group G, let sensitive value set be $\{s_1, \ldots, s_t\}$. s_j appears d_j times in G, j=1,...,t. Assume $d_1 \geq d_2 \ldots \geq d_t$. In

identity-reserved k-anonymity ($n \geq k$), we don't consider the distribution and background knowledge on sensitive attribute. So the probability of recognizing a_i is $c_i/\sum_{i=1}^{n} c_i$. If $c_1 = 1$, the probability is 1/n. That is the case of the common k-anonymity. If $c_1 \gg (c_2 + \ldots + c_n)$, the most of tuples in G correspond to a_1 and a_1 is easy to leakage of private information. This situation is similar to "homogeneity attack" discussed in [7].

In identity-reserved (k,l)-anonymity, we consider the diversity of sensitive attribute values. If we don't consider the distribution on ID attribute and other background knowledge, the probability of s_j is $d_j/\sum_{j=1}^{n} d_j$. These two principles simply take into account the diversity of ID or sensitive attribute, but they don't consider the frequency of these two attributes.

Identity-reserved (α,β)-anonymity confines the frequency of individual and sensitive value avoiding "homogeneity attack". If we only consider the identity or sensitive attribute respectively, the probability is no more than α or β.

3.4 Applicability

The identity-reserved anonymity takes the situation of multiple tuples per individual into account. We define the records per individual (rpi) of dataset to evaluate this situation, that is rpi=($the\ size\ of\ dataset$)/($the\ number\ of\ individuals$). If $rpi = 1$, each individual appears only once in the dataset. It's appropriate to use common anonymity. If $rpi > 1$, it's appropriate to use our anonymity for avoiding the privacy breach described before.

The identity-reserved anonymity holds the information between sensitive values of an individual that is discarded in common anonymity. The information is meaningful in researches, such as the market basket analysis or related diseases research.

4 Implementing

In common anonymity, generalization and suppression are the main approaches to meet the anonymity principles [8]. In fact generalization with suppression reduces the generalization height, but removal of the tuples also reduces the utility of the published table. In this paper, we also apply generalization to achieve identity-reserved anonymity.

Before generalization, we first recode an individually identifying attribute. Recoded individually identifying value is just a randomized numeric symbol to discriminate different individuals.

4.1 Global Recoding

Global recoding requires that all values of an attribute should generalize to the same domain level in the generalization hierarchy. For example, all values in Birth date are generalized to year and month in the format "mm/yyyy". The algorithm is similar to existing global-recoding algorithm in [4,7]. It makes use of

monotonicity property in generalization lattice space. The generalization doesn't stop until the result table meets the privacy requirement. If a certain number of suppression is allowed, the generalization processing finishes with suppressing. If suppression isn't allowed, suppression threshold is set as 0. Algorithm1 is a single-dimensional global recoding algorithm.

Algorithm 1: global recoding algorithm

Input : Table T, Suppression threshold S
Output: Published table PT
1. PT=the relation after recoding individually identifying attribute on T;
2. while (tuples that don't meet identity-reserved anonymity on PT>S) do
2.1 choose a QI attribute on PT;
2.2 generalize the chosen QI attribute on PT;
3. Remove the tuples that don't meet identity-reserved anonymity in PT;
4. return PT;

4.2 Local Recoding

Global recoding may generate excessive distortion to data set. Local recoding applies generalization on tuples not attributes. In local recoding, we adopt generalization without suppression. Wong et al. [12] proposed a top-down local-recoding algorithm. This approach first generalizes all tuples completely into one equivalence class. Then tuples are specialized in iterations while maintaining the anonymity principle. The process continues until specialization can't take place.

Table 3. A patient table needing publishing

Tuple-No	ID	Zip	Disease
1	1318	10085	Hypertension
2	1318	10085	Hyperlipemia
3	5072	10086	Diabetes
4	8634	10087	Heart
5	7437	10075	Hypertension
6	7437	10075	Diabetes
7	3582	10076	Heart
8	5629	10077	Flu
9	4713	10050	Heart

In this section, we propose a bottom-up approach. In our approach, we first check all tuples and mark the tuples that meet the requirement with group-id. Then we generalize a QI attribute on tuples without group-id in iterations. In every step of generalization, those tuples that meet the identity-reserved anonymity requirement are marked with a group-id. At last, a few tuples may be left without group-id, which are called "orphans". These orphans can't be grouped as a

group whatever they are generalized to (For example, 5 tuples are left without group-id while k=7). To group these orphans, we first move some tuples from other groups which could lend some tuples while maintaining the anonymity. If all other groups have no additional tuples to lend, we merge each orphan into a neighbor group and generalize them to form a QI group at last.

Table 4. A published table satisfying identity-reserved 2-anonymity

Group-No	ID	Zip	Disease
1	1318	1008*	Hypertension
1	1318	1008*	Hyperlipemia
1	8634	1008*	Heart
2	7437	1007*	Hypertension
2	7437	1007*	Diabetes
2	3582	1007*	Heart
2	5629	1007*	Flu
3	4713	100**	Heart
3	5072	100**	Diabetes

Table 5. A published table satisfying identity-reserved 3-anonymity

Group-No	ID	Zip	Disease
1	1318	100**	Hypertension
1	1318	100**	Hyperlipemia
1	8634	100**	Heart
1	5072	100**	Diabetes
1	4713	100**	Heart
2	7437	1007*	Hypertension
2	7437	1007*	Diabetes
2	3582	1007*	Heart
2	5629	1007*	flu

Let us illustrate with an example in Table 3. Suppose the QI contains only zipcode and ID is the recoded randomized number. The individual “1318” appears twice as tuple 1 and 2, and the individual “7437” appears twice as tuple 5 and 6. Other 5 individuals appear once in the table. We require identity-reserved k-anonymity and set k=2. First we check the table and find no tuples can be marked in a group. So zipcode attribute generalizes once (such as 1008*). Then tuple 1-4 can be marked with group-id 1, and tuple 5-8 can be marked with group-id 2. Now each group has 3 distinct individuals and 4 tuples. Tuple 9 is left, which is called ”orphan”. So we first search whether a group can lend some

tuples while maintaining the anonymity. If we only require identity-reserved k-anonymity and set k=2, we move a tuple (such as Tuple-No=3) to join the orphan and form the group 3. That result is showed on Table 4. If we require identity-reserved 3-anonymity (or identity-reserved (0.5,0.5)-anonymity), we can't move any tuple to join the orphan. So we could merge the orphan to a group(such as Group 1) and generalize tuples in that group. That result is showed on Table 5. Algorithm2 is a single-dimensional local recoding algorithm.

Algorithm 2: local recoding algorithm

Input : Table T
Output: Published table PT
1. PT=the relation after recoding individually identifying attribute on T;
2. Check and mark the tuples on PT which meet the identity-reserved anonymity;
3. While (tuples without marking the group-id on PT) >0 and (not generalize to the top of hierarchy) do
3.1 choose a QI attribute of PT;
3.2 generalize the chosen QI attribute for tuples without group-id on PT;
3.3 check and mark the tuples on PT which meet the identity-reserved anonymity;
4. if (tuples without marking the group-id on PT) >0 then
4.1 move tuples from other group;
4.2 check and mark;
5. if (tuples without marking the group-id on PT) >0 then
5.1 merge left tuples to other group;
6. return PT;

5 Experiments

In this section, we evaluate the identity-reserved anonymity principles in an experimental scenario, demonstrating practical applicability of the approaches. First we check the vulnerable QI group ratio in the situation of multiple tuples per individual. Then we evaluate the distortion ratio between the common k-anonymity and identity-reserved k-anonymity. At last we compare the global recoding and local recoding methods.

Experimental data come from the Adult database of UCI Machine Learning Repository [11]. The Adult database contains 45,222 tuples from US census data. We remove tuples with missing values. Since we check the identity-reserved anonymity effect, we add an attribute "Id-number". We fill in id-number so that a certain frequency of individuals appear several tuples. Description of other attributes is the same as [2].

We first choose 40,000 tuples and fill in distinct id-number. We partition these tuples to three disjoint subsets, called A, B, C. For each tuple in subset B, we duplicate it with the same id-number and QI values, and generate a different

sensitive value. So each individual in subset B corresponds to 2 tuples. For each tuple in subset C, we duplicate it twice with the same id-number and QI values, and generate a distinct sensitive value respectively. So each individual in subset C corresponds to 3 tuples. Subset A is directly added to the final relation. Thus we get $|A| + 2|B| + 3|C|$ tuples in the relation. According to the rpi definition in section 3.4, $rpi = (|A| + 2|B| + 3|C|)/(|A| + |B| + |C|) = Ratio_A + 2Ratio_B + 3Ratio_C$. We set test datasets with 4 different rpis according to Table 6.

Table 6. Description of test datasets size and rpi

rpi	A ratio	B ratio	C ratio	Dateset Size
1.2	0.85	0.1	0.05	48,000
1.4	0.70	0.2	0.10	56,000
1.6	0.55	0.3	0.15	64,000
1.8	0.40	0.4	0.20	72,000

First we check the vulnerable QI group ratio in the situation of multiple tuples per individual. We adopt common k-anonymity method by ignoring the id-number and get the anonymized table PT. In PT, we define the vulnerable group as the group which contains at least k tuples and at most (k-1) individuals, that is it meets the common k-anonymity but cannot meet identity-reserved k-anonymity. So the vulnerable group ratio is defined as (the number of vulnerable groups)/(the number of all groups). When k increase or rpi decreases, the number of individual in a QI group increases. So the vulnerable group ratio decreases. Fig 1 shows this trend.

Especially, some groups in PT only contain one individual. We call them single value group. Single value group only exists when k is no more than the maximum tuple number per individual. Fig 2 shows that single value group ratio decreases as k increase or rpi decreases.

We evaluate information loss of anonymized table in terms of distortion ratio. Distortion ratio is defined to describe the cost of recoding of the dataset. In [12], distortion ratio is equal to the distortion of generalized dataset divided by the distortion of the fully generalized dataset. The distortion of a value is the height of generalized value. The distortion of a tuple is the sum of its each attribute value generalization height. Let $height_i$ be the height of the i^{th} tuple. Let Height be the height of the fully generalized tuple. So the distortion ratio of dataset is defined as:

$$distortion\ ratio = \frac{\sum_{i=1}^{TupleCount} height_i}{TupleCount \times Height}$$

We compare the common k-anonymity and identity-reserved k-anonymity in Fig 3 (rpi=1.2) and Fig 4 (rpi=1.4). We notice that common k-anonymity achieves less distortion ratio, but the difference is slight.

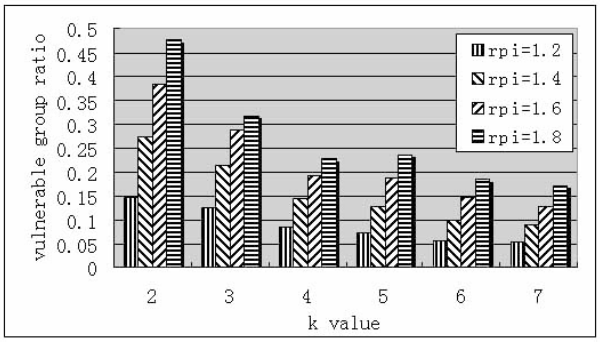

Fig. 1. Vulnerable group ratio with rpi and k

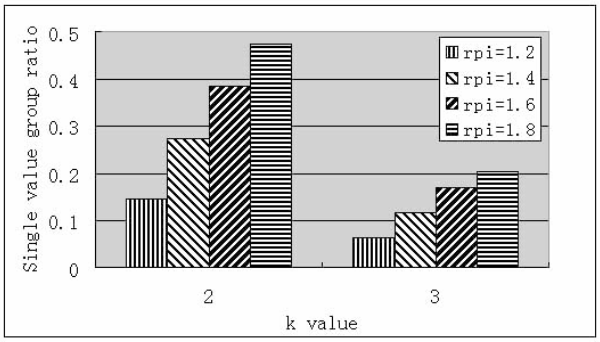

Fig. 2. Single value group ratio with rpi and k

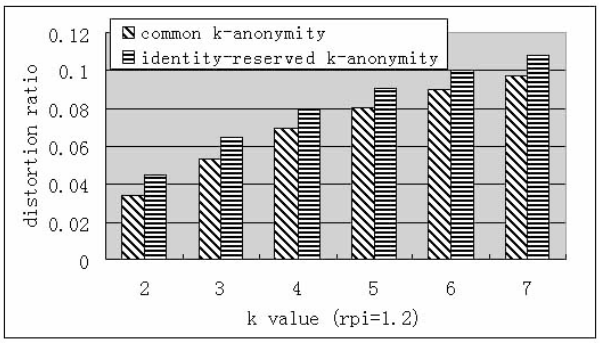

Fig. 3. Distortion ratio between common k-anonymity and identity-reserved k-anonymity when rpi=1.2

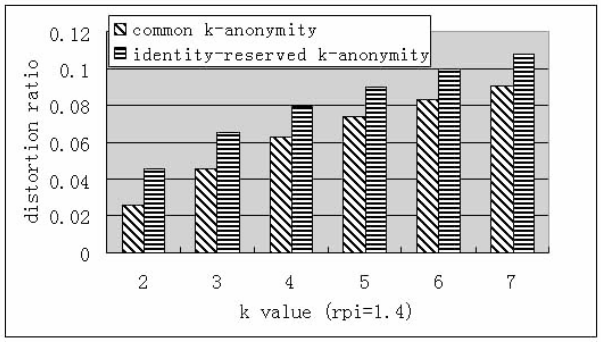

Fig. 4. Distortion ratio between common k-anonymity and identity-reserved k-anonymity when rpi=1.4

At last we compare local recoding and global recoding in identity-reserved anonymity. Figure 5 shows the distortion ratio of identity-reserved k-anonymity. When k increases, the distortion ratio increases slowly because more individuals have been generalized together. Since the global recoding algorithm generalizes the values to the same level on the hierarchy, the local recoding algorithm achieves much lower distortion. The global recoding algorithm with suppression achieves a bit lower distortion than that without suppression since it removes several outliers.

Figure 6 shows the distortion ratio of identity-reserved (k,l)-anonymity. In the experiments, l is usually less than k. When l increases with k fixed ($l \leq k$), the distortion ratio of global recoding keeps steady and that of local recoding increases slowly. The reason is that parameter l affects little compared to the parameter k when the number of all distinct sensitive values is larger than the parameter k and sensitive values distribute uniformly.

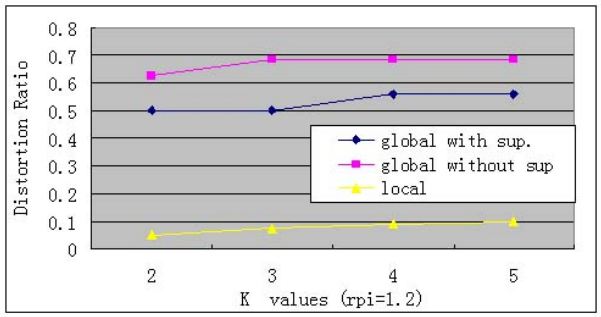

Fig. 5. Distortion ratio of identity-reserved k-anonymity

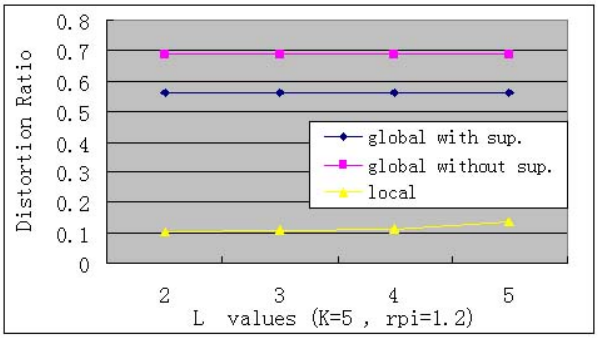

Fig. 6. Distortion ratio of identity-reserved (k,l)-anonymity

Figure 7 and 8 show the distortion ratio of identity-reserved (α,β)-anonymity using local recoding. In Fig 7, when β increases with α=0.5, distortion ratio decreases remarkably. When β is smaller, a QI group needs more distinct sensitive values. So it needs higher generalization level and distortion ratio. When β is large enough (such as 0.5) to match α, distortion ratio keeps steady because the table satisfying α usually satisfies that value of β at that time. In Fig 8, when α increases with β=0.5, distortion ratio decreases similar to Fig 7. Since the number of distinct identity values is much larger than that of sensitive value, the distortion ratio of Fig 7 decreases steeper than that of Fig 8.

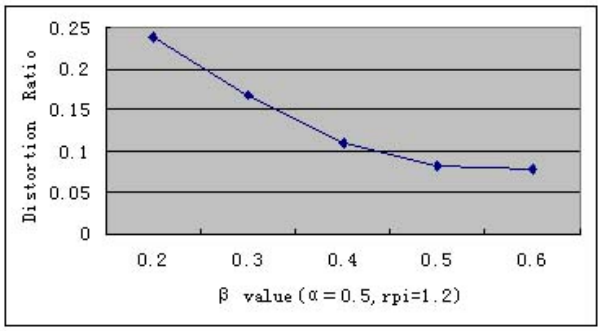

Fig. 7. Distortion ratio of identity-reserved (α, β)-anonymity

6 Conclusion

The current anonymity methods are inadequate since they can't take the individually identifying attribute into account. In this paper, we analyze the individually identifying attribute in the privacy preserving data publishing and pro-

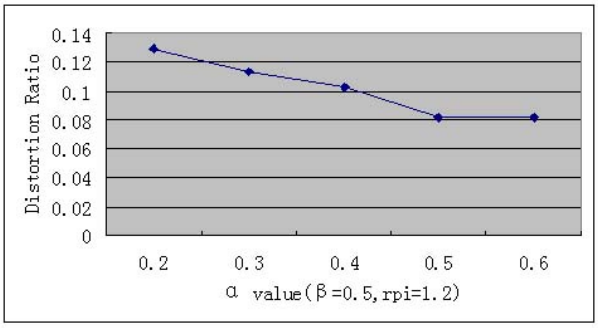

Fig. 8. Distortion ratio of identity-reserved (α, β)-anonymity

pose the concept of identity-reserved anonymity. We develop two approaches to achieve identity-reserved anonymity requirement. In local recoding, we propose a bottom-up algorithm which solves the orphan tuples by moving and merging. The algorithms are evaluated in an experimental scenario, demonstrating practical applicability of the approaches.

Acknowledgements

This work was supported by the National Science Foundation of China under Grant No.60403041. The authors would like to thank the anonymous reviewers for their insightful comments.

References

1. Aggarwal, G., Feder, T., Kenthapadi, K., Motwani, R., Panigrahy, R., Thomas, D., Zhu., A.:Anonymizing Tables. In *Proceedings of the 10th International Conference on Database Theory*, 246-258, 2005
2. Bayardo, R., Agrawal, R.: Data privacy through optimal k-anonymization. In *the 21st International Conference on Data Engineering*, 217-228, 2005
3. Fung, B. C. M., Wang, K., Yu, P. S.: Top-down Specialization for Information and Privacy Preservation. In *the 21st International Conference on Data Engineering*, 205-216,2005
4. LeFevre, K., DeWitt, D. J., Ramakrishnan, R.: Incognito: Efficient Full-domain K-anonymity. In *ACM International Conference on Management of Data*, 49-60, 2005
5. LeFevre, K., DeWitt, D. J., Ramakrishnan, R.: Mondrian Multidimensional K-Anonymity. In *the 22nd International Conference on Data Engineering*,25-35, 2006
6. Li, N.,Li, T.: t-Closeness: Privacy Beyond k-Anonymity and *l*-Diversity. In *the 23nd International Conference on Data Engineering*, 106-115, 2007
7. Machanavajjhala, A., Gehrke, J., Kifer, D.: *l*-diversity: Privacy beyond K-anonymity. In *the 22nd International Conference on Data Engineering*, 24-35, 2006

8. Samarati, P.: Protecting Respondents' Identities in Microdata Release. *IEEE Transactions on Knowledge and Data Engineering*, 13, 1010-1027, 2001
9. Sweeney, L.: Achieving K-anonymity Privacy Protection Using Generalization and Suppression. *International Journal on Uncertainty, Fuzziness and Knowledge Based Systems*, 10,571-588,2002
10. Sweeney, L.: K-anonymity: A Model for Protecting Privacy. *International Journal on Uncertainty, Fuzziness and Knowledge Based Systems*, 10, 557-570,2002
11. UCI Machine Learning Repository, http://www.ics.uci.edu/~ mlearn/MLRepository.html
12. Wong, R.C., Li, J., Fu, A.W., Wang, K.: (α,k)-anonymity: an Enhanced K-anonymity Model for Privacy-preserving Data Publishing. In *the 12th ACM SIGKDD*, 754-759, 2006
13. Xiao, X., Tao, Y.: Personalized Privacy Preservation. In *ACM International Conference on Management of Data*,229-240, 2006
14. Xiao, X., Tao, Y.: Anatomy: Simple and Effective Privacy Preservation. In *the 32nd international conference on Very large data bases*,139-150, 2006