On the Complexity of Restricted k-anonymity Problem*

Xiaoxun Sun[1] Hua Wang[1] and Jiuyong Li[2]

[1] Department of Mathematics & Computing
University of Southern Queensland, QLD, Australia
Email: {sunx, wang}@usq.edu.au
[2] School of Computer and Information Science
University of South Australia, Adelaide, Australia
Email: jiuyong.li@unisa.edu.au

Abstract. One of the emerging concepts in microdata protection is k-anonymity, introduced by Samarati and Sweeney. k-anonymity provides a simple and efficient approach to protect private individual information and is gaining increasing popularity. k-anonymity requires that every tuple(record) in the microdata table released be indistinguishably related to no fewer than k respondents. In this paper, we introduce two new variants of the k-anonymity problem, namely, the *Restricted k-anonymity problem* and *Restricted k-anonymity problem on attribute* (where suppressing the entire attribute is allowed). We prove that both problems are $\mathcal{NP}$-hard for $k \geq 3$. The results imply the main results obtained by Meyerson and Williams. On the positive side, we develop a polynomial time algorithm for the *Restricted 2-anonymity problem* by giving a graphical representation of the microdata table.

1 Introduction

Today's globally networked society places great demand on the sharing of information. However, the use of data containing personal information has to be restricted in order to protect individual privacy. To ensure the anonymity of the entities to which the sensitive data undergoing public or semipublic release refer, data holders often remove or encrypt explicit identifiers such as names, medical care card numbers(MCN) and addresses. The process is called de-identifying the data.

However, such a de-identification procedure does not guarantee the privacy of individuals in the data. Released information often contains other data, such as race, date of birth, gender and Zip code, which can be linked to publicly available information to re-identify respondents and

* This research was funded by Australian Research Council (ARC) grant DP0774450 titled "*Privacy Preserving Data Sharing in Data Mining Environments*".

to infer information that was not intended for release. Sweeney reported that 87 percent of the population of the United States can be uniquely identified by the combinations of attributes: gender, date of birth, and 5-digit zip code [11].

Besides de-identification, an alternative approach is to restrict the release of information in some way. In this paper, we focus on the strategy of k-anonymity, which was first proposed by Samarati and Sweeney [10]. A microdata table satisfies k-anonymity if every record in the table is identical to at least $(k-1)$ other records with respect to the set of quasi-identifier attributes.[1] Such a data set is called k-anonymous. As a result, an individual is indistinguishable from at least $(k-1)$ individuals in a k-anonymous data set.

Among the techniques proposed for providing anonymity in the release of microdata, the k-anonymity proposal focuses on two techniques in particular: generalization and suppression, which unlike other existing techniques, such as de-identification, preserve the truthfulness of the information. Generalization consists in substituting the values of a given attribute with more general values. We use $*$ to denote the more general value. For instance, we could generalize two different Zip code 4350 and 4373 to 435$*$. The other technique, referred to as data suppression, removes the part (cell suppression) or entire value (attribute suppression) of attributes from the microdata table. Note that suppressing an attribute to reach k-anonymity can equivalently be modeled via a generalization of all the attribute values to $*$.[2]

To illustrate the concept, consider the data in Table 1, which exemplifies medical data to be released after de-identification. This table does not contain personal identification attributes, such as name, address, and medical care card number(MCN). However, values of other released attributes, such as age, gender and Zip may appear in some external table jointly with the individual identity, and can therefore allow tracking. For instance, age, gender and Zip can be linked within Table 3 to reveal Name, Address, and City. In Table 1, for example, the first record is unique in these three attributes, and this combination, if unique in the external world as well, uniquely identifies the corresponding tuple as pertaining to

[1] The set of attributes included in the microdata table, also externally available and therefore exploitable for linking is called *quasi-identifier*.

[2] This observation holds assuming that attribute suppression removes only the values and not the attribute (column) itself. This assumption is reasonable since removal of the attribute (column) is not needed for k-anonymity.

MCN	Gender	Age	Zip	Diseases
∗	Male	25	4350	Hypertension
∗	Male	23	4351	Hypertension
∗	Male	22	4352	Depression
∗	Female	28	4353	Chest Pain
∗	Female	34	4352	Obesity
∗	Female	31	4350	Flu

Table 1: De-identified Private Table

MCN	Gender	Age	Zip	Diseases
∗	Male	22-25	435∗	Hypertension
∗	Male	22-25	435∗	Hypertension
∗	Male	22-25	435∗	Depression
∗	Female	28-34	435∗	Chest Pain
∗	Female	28-34	435∗	Obesity
∗	Female	28-34	435∗	Flu

Table 2: A 3-anonymous view of Table 1

"Lee, 10 Collard Court, Toowoomba", thus revealing that he has reported Hypertension.

Name	Address	City	Age	Zip	Gender
.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
Lee	10 Collard Court	Toowoomba	25	4350	Male
.....	.....	.....	.....	.....	.....

Table 3: Non de-identified Publicly available table

To avoid breaching privacy, Table 1 can be modified to Table 2. In Table 2, age is grouped into intervals, and Zips are clustered into large areas (the symbol $*$ denotes any digit). A (tuple) record in the quasi-identifier is identical to at least three other records in Table 2, and therefore, no individual is identifiable.

A k-anonymous table protects individual privacy in the sense that, even if an adversary has access to all the quasi-identifier attributes of all the individuals represented in the table, he would not be able to track down an individual's record further than a set of at least k records. Thus, releasing a k-anonymous table prevents definitive record linkages with publicly available databases and keeps each individual hidden in a crowd of $k-1$ other people.

In recent years, numerous algorithms have been proposed for implementing k-anonymity via generalization and suppression. Samarati [9] presents an algorithm that exploits a binary search on the domain generalization hierarchy to find minimal k-anonymous table. We recently improve his algorithm by integrating the hash-based technique [12]. Bayardo and Agrawal [3] presents an optimal algorithm that starts from a fully generalized table and specializes the dataset in a minimal k-anonymous table, exploiting ad hoc pruning techniques. LeFevre, DeWitt and Ra-

makrishnan [7] describes an algorithm that uses a bottom-up technique and a priori computation. Fung, Wang and Yu [4] present a top-down heuristic to make a table to be released k-anonymous. The approach applies to both continuous and categorical attributes. As far as k-anonymity problem is concerned, fewer theoretical results were obtained. The exceptions are Meyerson and Williams [8] and Aggarwal et al. [1, 2] proved the optimal k-anonymity is $\mathcal{NP}$-hard (based on the number of cells and number of attributes that are generalized and suppressed) and describe approximation algorithms for optimal k-anonymity.

2 Paper organization and contributions

In Section 3, we introduce two new variants of the k-anonymity problem, namely, the *Restricted k-anonymity problem* and *Restricted k-anonymity problem on attribute* and we discusses the connection between *Restricted k-anonymity problem* and general k-anonymity problem which stresses the significance of investigating this new class anonymity problem.

Our first contribution is the $\mathcal{NP}$-hardness proof the *Restricted k-anonymity problem* and the *Restricted k-anonymity problem on attribute*, and are presented in Section 4 and 5. The theoretical results for *Restricted k-anonymity problem* also provide an alternative $\mathcal{NP}$-hardness proof of general k-anonymity problem, which imply the main results obtained in [1, 2, 8].

The second contribution is presented in Section 6. Through a graphical representation of the microdata table, we develop a polynomial time algorithm for the *Restricted 2-anonymity problem*. Considering the connection between *Restricted k-anonymity problem* and general k-anonymity problem, we could develop another efficient algorithm for general k-anonymity problem as well. We will include this application part in a separate paper. Finally, conclusions and future work are given in Section 7.

3 Restricted *k*-anonymity problem

We consider degree-m tuples in the private database to be m-dimensional vectors v_i, drawn from $\sum^m$, where $\sum$ is a (finite) alphabet of possible values for attributes(columns). Thus, the private databases under consideration are formally represented as subsets $V \subseteq \sum^m$. Let $*$ be a symbol that is not in $\sum$.

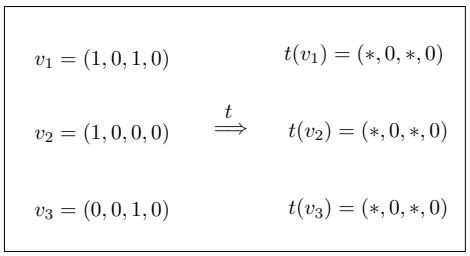

Fig. 1: Suppressing a dataset by t

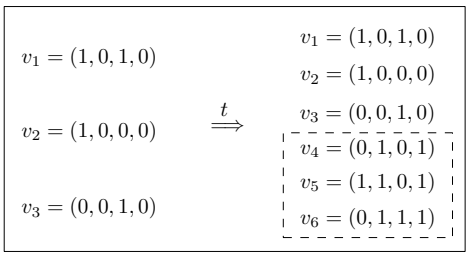

Fig. 2: Constructing a restricted instance

Definition 1. *Let t be a map from V to $(\sum \cup \{*\})^m$. We say that t is a suppressor on V if for all $v \in V$ and $j = 1, 2, \cdots, m$, it is the case that $t(v)[j] \in \{v[j], *\}$.*[3]

Intuitively speaking, a suppressor defines some kind of anonymous vector $t(v) = v'$ in an anonymous set $V' \subseteq (\sum \cup \{*\})^m$. The coordinates of V' are identical to the coordinates of V, except some may be suppressed by $*$. Consider following example. Let $V = \{1010, 1000, 0010\}$, with suppressor $t(a_1a_2a_3a_4) = *a_2 * a_4$ (each $a_i \in \{0, 1\}$), then the resulting $t(V) = \{*0 * 0, *0 * 0, *0 * 0\}$ (See Fig.1).

Now, we can extend the definition of a suppressor t to a set of vectors V. Here, we regard $t(V)$ as a multiset [4], when two or more vectors in V map to the same suppressed vector. (i.e. $v \neq v' \in V$, but $t(v) = t(v')$). Following, we define k-anonymity.

Definition 2. *Let t be a suppressor on the set $V = \{v_1, v_2, \cdots, v_n\} \subseteq \sum^m$. Then $t(V)$ is k-anonymous if and only if for all $v_i \in V$, there exists $k-1$ indices $i_1, i_2, \cdots, i_{k-1} \in \{1, 2, \cdots, n\}$, such that $t(v_{i_1}) = t(v_{i_2}) = \cdots = t(v_{i_{k-1}}) = t(v_{i_k})$.*

In other words, when a suppressor makes the database k-anonymous, it means that every anonymous vector is a member of a multiset of (at least) k identical vectors. For example, the left dataset in Fig.1 becomes 3-anonymous after suppressing by t.

Restricted k-anonymity problem: Given $V \subseteq \sum^m$ (where $\sum = \{0, 1\}$) such that the number of zeroes in each attribute (column) is exactly k; Is

[3] We consider a special case of suppressions, i.e. each entry is either included in the output, or omitted entirely, with a $*$ character taking its place.

[4] A multiset is a set in which elements can appear more than once. Notice that, given a multiset M and an element e, we may have that $e \in M$ match more than once; i.e., $\{e|e \in M\}$ is a multiset and its cardinality can be larger than 1. The usual set operations are extended to multisets accordingly.

there a suppressor t, such that $t(V)$ is k-anonymous and suppresses the minimum number of vector coordinates?

EXAMPLE: The left dataset in Fig.2 is an instance of general 3-anonymity problem and the right dataset is an instance of *Restricted 3-anonymity problem.*

Another version of the *Restricted k-anonymity problem* is where we choose whether or not to suppress various attributes from the database. We say that attribute(column) j is suppressed by t if for all $v \in V$, $v[j] = *$. Formally, we define the *Restricted k-anonymity problem on attribute* as follows:

Restricted k-anonymity problem on attribute: Given $V \subseteq \sum^m$ (where $\sum = \{0, 1\}$) such that the number of zeroes in each attribute (column) is exactly k; Is there a suppressor t, such that $t(V)$ is k-anonymous and suppresses the minimum number of attributes (columns)?

The reason why we introduce the definition of *Restricted k-anonymity problem* is its close connection with general k-anonymity problem. Given an instance of general k-anonymity problem, we could construct an instance of the *Restricted k-anonymity problem* (take Fig.2 as an example, by adding three vectors v_4, v_5, v_6, we can make the left dataset an instance of the *Restricted 3-anonymity problem*), which could provide an alternative approach to solve the general k-anonymity problem. Currently, we are working on developing algorithms to find solutions of the restricted problem and with some post-processing, the solution can be a good solution for the general problem.

4 Cell suppression is hard

In this section, we prove that the *Restricted k-anonymity problem* is $\mathcal{NP}$-hard for $k \geq 3$. First, recall the *Restricted 3-anonymity problem.*
Restricted 3-anonymity problem: Given $V = \{v_1, v_2, \cdots, v_n\} \subseteq \sum^m$ (where $\sum = \{0, 1\}$) such that the number of zeroes in each attribute (column) is 3, and $l \in N$: Is there a suppressor t, such that $t(V)$ is 3-anonymous, and the total number of suppressed vector coordinates in $t(V)$ is at most l?

Theorem 3. *The Restricted 3-anonymity problem is $\mathcal{NP}$-hard.*

Proof. The reduction is from *Exact cover by 3-sets (X3C)* [5]: Given a finite set X with $|X| = 3q$ and a collection C of 3-element subsets of X, does C contain an exact cover for X; that is, a sub collection $C' \subseteq C$ such that every element of X occurs in exactly one member of C'?

Let $X = (x_1, x_2, \cdots, x_{3q})$ and $C = (C_1, C_2, \cdots, C_m)$, where $|C_i| = 3$ for $i = 1, 2, \cdots, m$. We construct a database V as follows. For each x_i, define an m-dimensional vector $v_i \in \sum^m$:

$$v_i[j] = \begin{cases} 0 & \text{if } x_i \in C_j \\ 1 & \text{otherwise} \end{cases}$$

Set $V = (v_1, v_2, \cdots, v_{3q})$, $l = 3q(m-1)$, and the number of zeroes in each attribute(column) is 3 because of $|C_j| = 3$. Then V is an instance of the *Restricted 3-anonymity problem*.

Assume t is the optimal suppressor on V (i.e. suppresses the minimum number of vector coordinates and maintains 3-anonymity). We claim that the total number of coordinates suppressed by t is at most $3q(m-1)$ if and only if there is an *X3C* in C.

Sufficiency. Suppose that there is an *X3C* $C' \subseteq C$. For $i = 1, 2, \cdots, n$, let $j(i)$ be such that $C'_{j(i)}$ is the unique set in C' that contains x_i. Define a suppressor t by:

$$t(v_i)[j'] = \begin{cases} 0 & \text{if } j' = j(i) \\ * & \text{otherwise} \end{cases}$$

Since $x_i \in C'_{j(i)}$, the $v_i[j(i)] = 0$, and all the other are $*$. Therefore, t is a suppressor on V.

Now consider any $t(v_i)$. There are three elements x_i, $x_{i'}$, $x_{i''}$ in the set $C'_{j(i)}$, and each element has identical anonymous vectors; i.e. $t(v_i) = t(v_{i'}) = t(v_{i''})$. Hence there are two vectors in $t(V)$ which are identical to $t(v_i)$. This shows that $t(V)$ is 3-anonymous and t is feasible. Since in our solution, every $t(v) \in t(V)$ has exactly one non-$*$ coordinate, the number of $*$'s is exactly $3q(m-1)$. Therefore the optimal 3-anonymous solution has at most $3q(m-1)$ $*$'s in its vectors.

Necessity: Suppose that t suppresses at most $3q(m-1)$ coordinates and there does not exist *X3C*. We draw a contradiction as follows. Consider any 3-anonymous solution t for V. First, we answer the question: can there exist a vector with two non-$*$'s in its anonymous form? Suppose that v_i is such a vector. Since $t(V)$ is 3-anonymous, there must exist two other vectors $v_{i'}$ and $v_{i''}$, which have the same value as $t(v_i)$ in the anonymous form, say, $t(v_{i'})$ and $t(v_{i''})$. Since the non-$*$ coordinates have the same values as in the original v_i vectors, we must have v_i, $v_{i'}$ and $v_{i''}$ identical in two different coordinates, j and j'. By construction, any two vectors in V can match only in coordinates where they are 0, and $v_i[j] = 0$ only if the element x_i is in the set C_j . Hence v_i, $v_{i'}$ and $v_{i''}$ are in the two

different sets, C_j and $C_{j'}$ of C. However, this means two different sets are identical in C, which is not possible. So for any feasible 3-anonymous suppressor t for V, every vector $v_i \in V$ has at most one non-$*$ coordinate in its 3-anonymous form $t(v_i)$. Hence at least $3q(m-1)$ coordinates in $t(V)$ are suppressed.

Therefore, if we have a $t(V)$ with at most $3q(m-1)$ suppressed coordinates, it must be that every vector in $t(V)$ has exactly one non-$*$ coordinate. Given this fact, we can construct an *X3C* C' for C in the following way. For each $i = 1, 2, \cdots, n$, consider the non-$*$ coordinate in $t(v_i)$. This coordinate must have value 0 (otherwise there can be no identical vectors). If this corresponds to the coordinate j, we add the set C_j to a cover C'. Clearly we produce a collection of sets such that each element in X is in at least one set. Since there are 3 identical vectors for every vector $v \in V$(including v), it follows that there are at most q sets in C'. Since we need at least q sets to cover every element, there must be exactly q sets in C', which is exactly an *X3C*. This contradicts our assumption, so it follows that there is an *X3C* in C if and only if the optimal 3-anonymous solution has at most $3q(m-1)*$'s.

Corollary 4. *The Restricted k-anonymity problem is $\mathcal{NP}$-hard for $k \geq 3$.*

Corollary 5. *The k-anonymity problem is $\mathcal{NP}$-hard for $k \geq 3$.*

Corollary 5 was first obtained by Meyerson and Williams [8].

5 Attribute suppression is hard

In this section, we consider the situation whether or not to suppress various attributes from the database and we prove it is hard as well.

Suppose that $X = (x_1, x_2, \cdots, x_{3q})$ and $C = (C_1, C_2, \cdots, C_m)$, where $|C_i| = 3$, for $i = 1, 2, \cdots, m$, and let $\sum = \{0, 1\}$. We build the database $V = (v_1, v_2, \cdots, v_{3q})$ where each v_i represents an element in X. Assume t suppresses the least number of attributes and is defined as in Theorem 3.

Theorem 6. *The Restricted 3-anonymity problem on attribute is $\mathcal{NP}$-hard.*

Proof. (proof sketch) We claim that there exists a suppressor that suppresses at most $m - q$ attributes and maintains 3-anonymous if and only if C has an *X3C*. If C has an *X3C*, then by suppressing those $m - q$ attributes not in the cover, each remaining attributes has 3 vectors that has

the same value, which is 3-anonymous. Conversely, if we have a suppressor t as above, then for every j, since the anonymous table is 3-anonymous and the number of zeroes is 3, there are exactly 3 vectors v_k such that $v_k[j] = 0$. It follows that if an attribute is not suppressed, then there exists 3 vectors with the same value under this attribute. Since the two attributes i and j are not suppressed in a 3-anonymous table if and only if $C_i \cap C_j = \emptyset$, at least $m - q$ attributes must be suppressed in any 3-anonymous table. Therefore, if we obtain a $t(V)$ with at most $m-q$ suppressed attributes, it must be that exactly $m-q$ attributes are suppressed in a 3-anonymous table. Then we can obtain an *X3C* as in Theorem 3.

Corollary 7. *The Restricted k-anonymity problem on attribute is $\mathcal{NP}$-hard for $k \geq 3$.*

Corollary 8. *The k-anonymity problem on attribute is $\mathcal{NP}$-hard for $k \geq 3$.*

Corollary 8 implies the result obtained by Aggarwal *et al.*[1, 2].

6 Algorithm for Restricted 2-anonymity problem

In this section, we present a graphic representation of the *Restricted 2-anonymity problem*, which produces a polynomial time algorithm with running time in $O(n^2m)$. First, recall the *Restricted 2-anonymity problem*:

PROBLEM: *Restricted* 2*-anonymity problem*
INSTANCE: Dataset $V = \{v_1, v_2, \cdots, v_n\} \subseteq \sum^m$, where $\sum = \{0, 1\}$ and the number of zeroes in each attribute (column) is 2, and $l \in N$.
QUESTION: Is there a suppressor t such that $t(V)$ is 2-anonymous and the total number of suppressed vector coordinates in $t(V)$ is at most l?

The transformation is made from the *perfect matching problem* in a simple graph. We include its definition here for completeness.
Perfect matching problem: Given a graph $G = (U, E)$ with $|U| = n$ and $|E| = m$, is there a subset $S \in E$ of $n/2$ edges such that each vertex of U is contained in exactly one edge of S?

Without loss of generality, assume that no two columns in the dataset have the same values. (If not, we could simplify the dataset by deleting the repeated one, which has no effect on the anonymity process) Also assume that $V = \{v_1, v_2, \cdots, v_n\} \in \sum^m$. Now construct a graph as follows:

Let $U = (v_1, v_2, \cdots, v_n)$ and $E = \{e_{ik}(j)\}$ where for each $j = 1, 2, \cdots, m$, $e_{ik} = (v_i(j), v_k(j))$ with $v_i(j) = v_k(j) = 0$ according to the assumption.

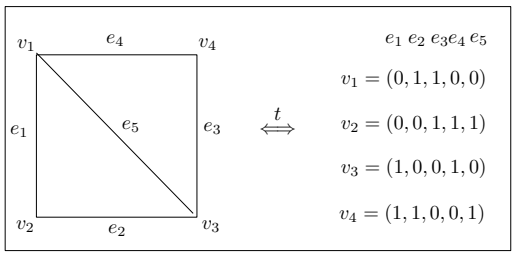

Fig. 3: Dataset(right)and its graphical representation(left)

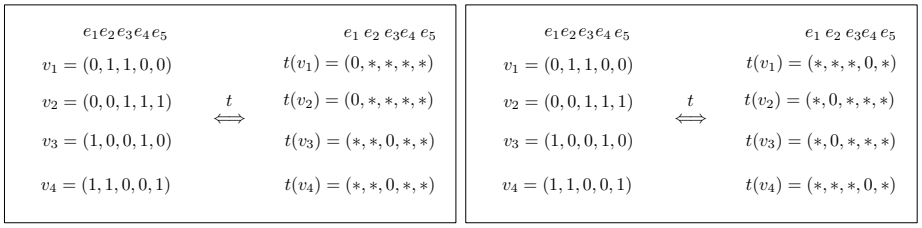

Fig. 4: Dataset and its two 2-anonymous tables

Then we get the simple graph $G = (U, E)$ with $|U| = n$ and $|E| = m$. (See Fig.3 as an example.) On the contrary, if we have the simple graph $G = (U, E)$ with $U = (v_1, v_2, \cdots, v_n)$ and $E = (e_1, e_2, \cdots, e_m)$, then construct a database V as follows:

For each v_i, define an m-dimensional vector $v_i \in \sum^m$ as $v_i[j] = 0$ if $v_i \in e_j$; Otherwise, $v_i[j] = 1$; Set $V = \{v_1, v_2, \cdots, v_n\}$. Then because the graph G is simple, obviously, the number of zeroes in each attribute (column) is 2, which is an instance of the *Restricted* 2*-anonymity problem.*

Theorem 9. *Given an instance of the Restricted 2-anonymity problem, the optimal Restricted 2-anonymous solution has at most* $n(m-1)$ $*$*'s suppressed by* t *if and only if there is a perfect matching in the corresponding constructed graph* G*.*

Proof. The proof is similar to Theorem 3, we omit it here due to the page limit.

Corollary 10. *The Restricted 2-anonymity problem can be solved in polynomial time.*

Algorithm 1: Polynomial time algorithm for the *Restricted 2-anonymity problem.*

Input : A dataset $V = (v_1, v_2, \cdots, v_n) \subseteq \sum^m$

Output: The 2-anonymous dataset $t(V)$ (where t is a suppressor)

1. Construct the graph $G = (U, E)$ where $U = (v_1, v_2, \cdots, v_n)$ and $E = \{e_{ik}(j)\}$ and for each $j = 1, 2, \cdots, m$, $e_{ik} = (v_i(j), v_k(j))$,with $v_i(j) = v_k(j) = 0$
2. Find one perfect matching M in G.
3. If found, let $M(i)$ denote the unique edge in M containing node i and let $t(M(i)) = 0$ and $t(j) = *$, if $j \neq M(i)$. Output $t(V)$.
4. If not found. Output $t(V)$ with each value replaced by $*$ in V.

Running Time: The running time of Algorithm 1 depends on Step 2, which can be solved in $O(n^2m)$ [6]. Since the transformation could be done in at most $O(n^2)$, the algorithm time complexity for *Restricted 2-anonymity problem* is in $O(n^2m)$. Also, since the graph can be specified by its vertex adjacency matrix A, which could be described by at most nm bits of input, so the space (memory) complexity of the algorithm is $O(nm)$. Note that if we find out all the perfect matchings M in G, then we could find all the possible 2-anonymous tables.

EXAMPLE: We use Fig.4 as an example to illustrate how Algorithm 1 works. Our objective is to make the left dataset in Fig.4 2-anonymous. The left graph in Fig.3 is the graphic representation of the dataset in Fig.4. In that graph, we could find all perfect matchings $\{e_1, e_3\}$ and $\{e_2, e_4\}$ and according to Algorithm 1, all the 2-anonymous tables are shown in Fig.4.

7 Conclusions and future work

In this paper, we introduce two new variants of the k-anonymity problem, namely, the *Restricted k-anonymity problem* and *Restricted k-anonymity problem on attribute*. We prove that both problems are $\mathcal{NP}$-hard for $k \geq 3$. The results imply the main results obtained by Meyerson and Williams. We have also developed a polynomial time algorithm for the *Restricted 2-anonymity problem* by giving a graphical representation of the microdata table.

Our future work is to develop applicable algorithms for general k-anonymity problem based on the theoretical results obtained in this paper. More specifically, it involves developing a new efficient exact algorithm and providing better approximate algorithm scheme for general k-anonymity problem based on the connection and transformation between the *Restricted k-anonymity problem* and general k-anonymity problem.

Acknowledgements

We would like to thank Professor Jeffrey Yu for his useful comments on the paper.

References

1. Aggarwal G, Feder T, Kenthapadi K, Motwani R, Panigrahy R, Thomas D, Zhu A. Anonymizing tables. In *Proc. of the 10th International Conference on Database Theory (ICDT05)*, pp. 246-258, Edinburgh, Scotland.
2. Aggarwal G, Feder T, Kenthapadi K, Motwani R, Panigrahy R, Thomas D, Zhu A. Approximation algorithms for k-anonymity. *Journal of Privacy Technology*, paper number 20051120001.
3. Bayardo R J, Agrawal R. Data privacy through optimal k-anonymization. *In Proc. of the 21st International Conference on Data Engineering (ICDE05)*, pp. 217-228, Tokyo, Japan.
4. Fung B, Wang K, Yu P. Top-down specialization for information and privacy preservation. *In Proc. of the 21st International Conference on Data Engineering (ICDE05)*, Tokyo, Japan.
5. Garey M R, Johnson D S. *Computers and Intractability: A Guide to the Theory of $\mathcal{NP}$-Completeness*. San Francisco. Freeman, 1979
6. Lawler E L. *Combinatorial Optimization: Networks and matroids*. Holt, Rinehart and Winston, New York, 1976
7. LeFevre K, DeWitt DJ, Ramakrishnan R. Incognito: Efficient fulldomain k-anonymity. *In Proc. of the 24th ACM SIGMOD International Conference on Management of Data*, pp. 49-60, Baltimore, Maryland, USA, 2005.
8. Meyerson A, Williams R. On the complexity of optimal k-anonymity. *In Proc. of the 23rd ACM-SIGMOD-SIGACT-SIGART Symposium on the Principles of Database Systems*, pp. 223-228, Paris, France, 2004.
9. Samarati P. Protecting respondents' identities in microdata release. *IEEE Transactions on Knowledge and Data Engineering*, 13(6):1010-1027. 2001
10. Samarati P, Sweeney L. Generalizing Data to Provide Anonymity when Disclosing Information (Abstract). *In Proc. of ACM Symposium on Principles of Database Systems*, pp. 188, 1998.
11. Sweeney L, k-anonymity: a model for protecting privacy. *International journal on uncertainty, Fuzziness and knowledge based systems*, 10(5):557-570, 2002.
12. Sun X, Li M, Wang H and Plank A. An efficient hash-based algorithm for minimal k-anonymity problem. *to appear in Thirty-First Australasian Computer Science Conference (ACSC2008)*, Wollongong, Australia.