Optimized data de-identification using multidimensional k-anonymity

Kai-Cheng Liu Chuan-Wei Kuo* Wen-Chiuan Liao Pang-Chieh Wang
Computational Intelligence Technology Center (CITC)
Industrial Technology Research Institute (ITRI)
Hsinchu, Taiwan
{AustinLiu, auberonkuo, liao, pangchieh}@itri.org.tw

Abstract—**In the globalized knowledge economy, big data analytics have been widely applied in diverse areas. A critical issue in big data analysis on personal information is the possible leak of personal privacy. Therefore, it is necessary to have an anonymization-based de-identification method to avoid undesirable privacy leak. Such method can prevent published data form being traced back to personal privacy. Prior empirical researches have provided approaches to reduce privacy leak risk, e.g. Maximum Distance to Average Vector (MDAV), Condensation Approach and Differential Privacy. However, previous methods inevitably generate synthetic data of different sizes and is thus unsuitable for general use. To satisfy the need of general use, k-anonymity can be chosen as a privacy protection mechanism in the de-identification process to ensure the data not to be distorted, because k-anonymity is strong in both protecting privacy and preserving data authenticity. Accordingly, this study proposes an optimized multidimensional method for anonymizing data based on both the priority weight-adjusted method and the mean difference recommending tree method (MDR tree method). The results of this study reveal that this new method generate more reliable anonymous data and reduce the information loss rate.**

Keywords—privacy preserving; k-anonymity; de-identification; data quality; information loss

I. INTRODUCTION

The field of privacy preservation has undergone many fluctuations and shifts with the increasing use of data mining, machine learning, and other analysis applications supported by big data. When a dataset containing individuals' private data is released to others, it must be anonymized to reduce the risk of re-identification [1]–[6]. In real cases, the k-anonymity privacy preservation mechanism has been used widely owing to its simplicity and efficiency[7]–[9]. The concept of k-anonymity was first proposed in 1998 [7], and a few k-anonymity based algorithms have been developed subsequently [9]–[12]. Descriptions of the related definitions of k-anonymity are given below:

In a k-anonymity privacy preservation mechanism, there is a table $\mathrm{T} = (\mathrm{A}_1, \dots, \mathrm{A}_\mathrm{n}), \mathrm{n} \in \mathrm{N}$. Each attribute $A_k, 1 \leq k \leq n$ corresponds to one single data type, and the various types of data are as follows:

(1) $\mathrm{Identifiers} = (\mathrm{A}_\alpha, \dots, \mathrm{A}_\beta), 1 \leq \alpha \leq \beta \leq \mathrm{n}$
(2) $QuasiIdentifiers = (A_i, \dots, A_j), 1 \leq i \leq j \leq n$
(3) $Sensitives = (A_x, \dots, A_y), 1 \leq x \leq y \leq n$
(4) $Insensitives = (A_w, \dots, A_m), 1 \leq w \leq m \leq n$
$(\ \forall \alpha \neq \beta \neq i \neq j \neq x \neq y \neq w \neq m\ )$

In the beginning, the privacy preservation algorithms focused only on using a single-dimensional way to achieve k-anonymity privacy protection (i.e., they generalize one column during one operation) [13], [14]. Although the single-dimensional method usually outputs data with a low information loss rate [15], which maintains data quality, the method is very resource-intensive, and it requires a lot of time to output the result, especially when dealing with big data. With the increasing usage of large data applications (e.g., data mining or machine learning), a few studies have been carried out on the effectiveness of using multidimensional methods (i.e., where all columns are generalized during an operation) to generate the de-identifying process [16], [17]. A framework called basic Mondrian algorithm for multidimensional k-anonymity-based algorithms that offers a sound theoretical basis for developing multidimensional k-anonymity models has been proposed is a previous study [17]. However, this multidimensional k-anonymity method of de-identification cannot maintain the analytical nature of data, and the result is usually unusable for analysis. In this paper, we propose an optimized multidimensional k-anonymity model, which ensures that the de-identification process outputs data with the characteristics: 1) Quality of data is maintained. 2) The output is suitable for general analytical usage. Figure 1 shows the framework of our optimized multidimensional k-anonymity model. This model was generated using a top-down greedy algorithm, and we used two particular methods in this greedy algorithm, namely, priority weight-adjusted method and the MDR tree method, to enhance the quality of the output data and make them suitable for analysis.

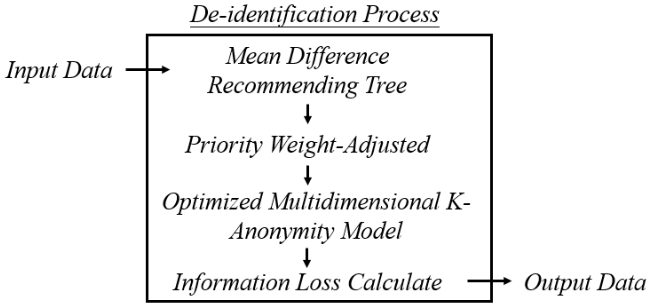

De-identification Process
Input Data
Mean Difference Recommending Tree
Priority Weight-Adjusted
Optimized Multidimensional K-Anonymity Model
Information Loss Calculate
Output Data

Figure 1. Framework of optimized multidimensional k-anonymity model.

The remainder of this paper is organized as follows:

A three-phase study is designed to generate the optimized multidimensional k-anonymity model. The first section deals with the theoretical foundations for developing the research. In sections 2 and 3, the research methodology is presented, with details of the model used in the research, algorithm, and procedures. In section 4, a comparison between the results and methodologies is presented, along with a thorough description of the operation process of the optimized multidimensional k-anonymity model. Finally, in section 5, the conclusions are presented.

II. Background and Related Work

Previous empirical studies have been crucial in laying the foundation for generating the optimized multidimensional k-anonymity model, which include a highly accurate information loss metrics model and a basic multidimensional k-anonymity algorithm [16].

Calculating the information loss rate of output data after a privacy preservation process is very important[18], [19]. We used the weighted normalized penalty model for determining information loss metrics [15]. More specifically, information loss is inevitable when a privacy preservation procedure is applied to raw data. Therefore, we need a suitable methodology to measure the information loss rate. A mathematical model was proposed in a previous study [15] for determining the value of information loss metrics as follows.

A. Normalized Certainty Penalty Model of Categorical Attributes

Compared to numerical attributes, categorical attributes need to consider the attribute hierarchy. Let a hierarchy tree including a set of leaf nodes $C_1, \dots, C_m$, and let U be a node in the hierarchy tree which is an ancestor of $C_1, \dots, C_m$. The size of U is the number of leaf nodes that are descendants of U, denoted as size (U). Consider a table T with a categorical quasi-identifier A_γ, where the value of A_γ is C. The value C will be replaced by a set of values $C_1, \dots, C_m$ when generalizing for anonymization. Here, $C_1, \dots, C_m$ are from the same generalized group of attribute A_γ. The normalized certainty penalty of t is given as follows:

$$\mathrm{NCP}_{\mathrm{A}_\gamma}(\mathrm{t})_{\mathrm{t}\epsilon\mathrm{T}} \ \frac{\mathrm{size(U)}}{|\mathrm{A}_\gamma|},$$

where $|\mathrm{A}_\gamma|$ is the number of distinct values of A_γ

The sum of the weighted normalized certainty penalty of all tuples is the total weighted normalized certainty penalty as follows:

$$NCP(T) = \sum_{t \in T} \sum_{\gamma=1}^{m} (w_\gamma \cdot NCP_{A_\gamma}(t))$$

B. Basic Multidimensional K-Anonymity Algorithm

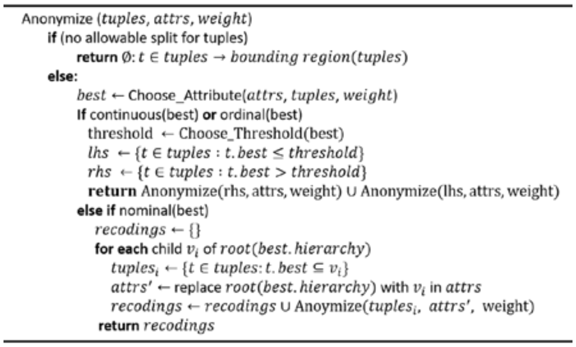
The detailed definition of single-dimensional k-anonymity is given as follows:

Consider a table T that includes the quasi-identifiers of n attributes $(Q_1, \dots, Q_n), n \in N$. Each attribute $\mathrm{Q_k}$ has a domain of values $\mathrm{M_{Q_k}}$. Single-dimensional recoding (single-dimensional way k-anonymity) is then given as a set of functions $\theta_1, \dots, \theta_n$, and each function $\theta_k\colon \mathrm{M_{Q_k}} \rightarrow \mathrm{M}^{'}$.

The detailed definition of multidimensional k-anonymity is as follows:

Consider a table T that includes the quasi-identifiers of n attributes $(Q_1, \dots, Q_n), n \in N$. Each attribute $\mathrm{Q_k}$ has a domain of values $\mathrm{M_{Q_k}}$. Multidimensional recoding (multidimensional k-anonymity) is then given as a function θ, and the function $\theta\colon M_{Q_1} \times M_{Q_2} \times \dots \times M_{Q_k} \rightarrow M'$.

Owing to its computational limitations, the single-dimensional k-anonymity algorithm cannot cope with the increasing amounts of data generated recently (i.e., calculation time is too long to complete all privacy preservation processes within limited time). [17] provided an useful theoretical framework to use the multidimensional k-anonymity model to solve this problem. The algorithm is shown in Figure 2.



```
Anonymize ($tuples, attrs, weight$)
    if (no allowable split for tuples)
        return $\emptyset: t \in tuples \rightarrow bounding\ region(tuples)$
    else:
        $best \leftarrow$ Choose_Attribute($attrs, tuples, weight$)
        If continuous(best) or ordinal(best)
          threshold $\leftarrow$ Choose_Threshold(best)
          $lhs \leftarrow \{t \in tuples : t.best \leq threshold\}$
          $rhs \leftarrow \{t \in tuples : t.best > threshold\}$
          return Anonymize(rhs, attrs, weight) $\cup$ Anonymize(lhs, attrs, weight)
        else if nominal(best)
           $recodings \leftarrow \{\}$
           for each child $v_i$ of $root(best.hierarchy)$
              $tuples_i \leftarrow \{t \in tuples: t.best \subseteq v_i\}$
              $attrs' \leftarrow$ replace $root(best.hierarchy)$ with $v_i$ in $attrs$
              $recodings \leftarrow recodings \cup$ Anoymize($tuples_i$, $attrs'$, weight)
            return $recodings$
```

Figure 2. Basic Mondrian algorithm

This algorithm uses a top-down greedy algorithm structure. First, the information loss rate is computed for each attribute in the table. Second, based on the order of the information loss rate of each attribute, the algorithm decides the attributes that will be used to conduct de-identification and then executes the de-identification process for each tuple. Finally, the algorithm repeats itself from the first step until there are no tuples that can be split further. We generated an optimized multidimensional k-anonymity model based on this framework.

III. Optimized Multidimensional K-Anonymity Model

Even though the basic Mondrian algorithm has the undeniable merit of offering valuable insights into privacy preservation, it has some limitations [17]. The quality of output data must be enhanced to satisfy analytical needs. Moreover, the information loss rate must be decreased. In this section, we describe the designs of two particular methods for improving the basic Mondrian algorithm to generate an optimized multidimensional k-anonymity model as a privacy preservation method.

A. MDR Tree Method

[16] provided the Mondrian algorithm to generalize the numerical attributes in one table and then provided the basic Mondrian algorithm to generalize categorical attributes [17]. Assume a table T that includes categorical attributes $\mathrm{A}_1, \ldots, \mathrm{A}_\mathrm{n}$, and each tuple t in table T has the attributes $a_1, \ldots, a_n$. In the basic Mondrian algorithm, each categorical attribute $A_1, \ldots, A_n$ generates a hierarchy tree. Figure 3 depicts metaphorically what happens in these hierarchy trees. The notion is that generating a hierarchy tree for categorical attributes has significant implications in terms of maintaining the analytical nature of data. As shown in Figure 3, basic Mondrian algorithm uses a hierarchy tree to distinguish the countries. There are three levels in this hierarchy tree, namely, level zero (*), level one (N. AMERICA, ASIA), and level two (the countries).

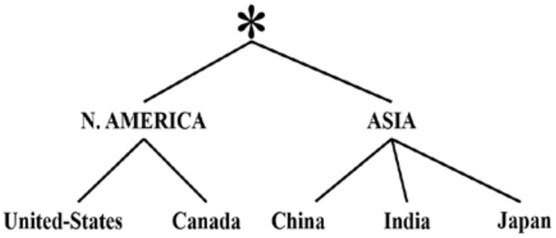


Figure 3. Example of categorical hierarchy tree

The basic Mondrian algorithm follows this hierarchy tree to generalize the categorical attributes. Overall, the results have been very positive for generalization of the categorical attributes. However, for numerical attributes, the Mondrian algorithm uses a partitioning method by splitting the numerical attributes with medians to complete privacy preservation. The results obtained with the numerical attributes are mixed. Each tuple may have different intervals for the same numerical attribute. Although the algorithm can reduce the information loss rate, the output data can barely be used for analysis owing to their irregular intervals.

The method described here could serve as a basis for dealing with numerical attributes. We propose a MDR tree for numerical attributes. It distinguishes the numerical attributes across four levels, from level zero to level three. For example, a table T has a numerical attribute A containing age data (age between 0 and 100 years). We divide the age into four levels: Level three means the data would be evenly partitioned into 20 groups. Each group has 5% of the total data. Level two means the data would be evenly partitioned into ten groups. Level one means the data would be evenly partitioned into five groups. Level zero covers the whole data. Moreover, for real data, the range of numerical attributes would not be fixed. Therefore, we must use the mean difference as the standard interval range (details are given in section 4).

The mean difference is computed as follows:

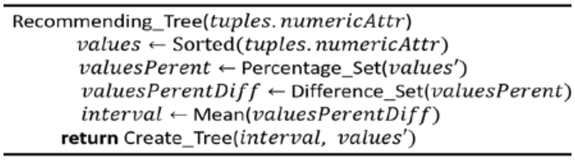
Assume we have a table T that includes a numerical attribute A, which has N tuples. Each tuple $\mathrm{t_i}, \forall \mathrm{t} \in \mathrm{T}, 1 \leq \mathrm{i} \leq \mathrm{N}$ has the attribute $a_i, 1 \leq i \leq N$of A. We first sort attribute A in ascending order to obtain a sorted attribute A'. For $\mathrm{A}' = (\mathrm{a}_1', \ldots, \mathrm{a}_\mathrm{N}')$, we observe and divide it by a 5% interval, where the interval is $\gamma_1, \ldots, \gamma_{20}$ (i.e., 0% to 100% in steps of 5%). The range of each interval is $M_1, \ldots, M_{20}$. Furthermore, a new interval MD of the numerical attribute A of a level three MDR tree is given as follows:

$$Mean\ Difference(MD) = \frac{\sum_{k=1}^{20} M_k}{20}$$

Finally, we re-segment the attribute A by using the interval MD, which means that we generate a new set of upper and lower bounds $\mathrm{S} = (\mathrm{s}_1, \ldots, \mathrm{s}_\mathrm{g}), \forall 1 \leq \mathrm{g} \leq \mathrm{U}$. Moreover, each tuple $t_i, \forall t \in T, 1 \leq i \leq N$ is re-assigned to the $s_g, 1 \leq g \leq U$ to which it belongs, and U refers to the number of S and is denoted as follows:

$$U = \frac{(\text{Maximum A}) - (\text{Minimum A})}{\text{MD}}$$

Figure 4 shows the entire process of this MDR tree method. In brief, we deal with the numerical attributes in a categorical fashion to ensure the output data can be used for analysis purposes.



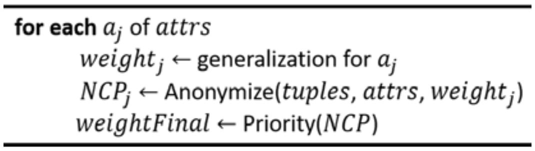
```
Recommending_Tree($tuples.numericAttr$)
        $values \leftarrow$ Sorted($tuples.numericAttr$)
        $valuesPerent \leftarrow$ Percentage_Set($values'$)
        $valuesPerentDiff \leftarrow$ Difference_Set($valuesPerent$)
        $interval \leftarrow$ Mean($valuesPerentDiff$)
    return Create_Tree($interval,\ values'$)
```

Figure 4. MDR tree algorithm

B. Priority Weight-Adjusted Method

As pointed out in section 2, the weight of an attribute serves an important role in the information loss metric. In this work, a priority weight-adjusted algorithm was developed to identify and determine the weights of the attributes that can influence the information loss rate of the optimized multidimensional k-anonymity model (see Figure 5). The priority weight-adjusted algorithm is shown as follows:

Suppose we have a table T containing the attributes $A_1, \ldots, A_n$, and there are N tuples. Each tuple t has attributes $a_1, \ldots, a_n$. We first execute the basic Mondrian algorithm by using the given attributes $A_i, \forall 1 \leq i \leq n$ having weight N, while the other attributes are assigned a weight of 1. This implies that the algorithm can first split this attribute. Until there are no tuples that can be split further at this attribute, the remainder of attributes would be considered. By executing this process, we can understand the influence of each attribute on NCP. That is, we run the basic Mondrian algorithm n times with the respective NCP. Then, we sort these n times of NCP to determine the priority. Each attribute A_i has its NCP_i, and the algorithm uses this result to determine the priority of the attributes (i.e., it will determine which attribute will be generalized first).

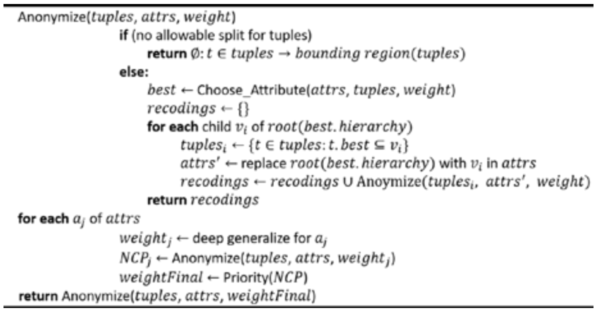


```
for each $a_j$ of $attrs$
        $weight_j \leftarrow$ generalization for $a_j$
        $NCP_j \leftarrow$ Anonymize($tuples, attrs, weight_j$)
        $weightFinal \leftarrow$ Priority($NCP$)
```

Figure 5. Priority weight-adjusted algorithm

C. Optimized Multidimensional K-Anonymity Model

The method of the optimized multidimensional k-anonymity model is based on two methods and one model that we have mentioned before, namely, basic Mondrian algorithm, priority weight-adjusted method, and MDR tree method. When running this model, the algorithm first generates an MDR tree for the numerical quasi-identifier attribute. Second, the priority weight-adjusted method is used to provide a set of weights to the attributes for use in the basic Mondrian algorithm. Finally, the model executes the basic Mondrian algorithm based on previous methods. Details of this model are shown schematically in Figure 6.



```
Anonymize($tuples, attrs, weight$)
        if (no allowable split for tuples)
            return $\emptyset: t \in tuples \rightarrow bounding\ region(tuples)$
        else:
            $best \leftarrow$ Choose_Attribute($attrs, tuples, weight$)
            $recodings \leftarrow \{\}$
            for each child $v_i$ of $root(best.hierarchy)$
                $tuples_i \leftarrow \{t \in tuples: t.best \subseteq v_i\}$
                $attrs' \leftarrow$ replace $root(best.hierarchy)$ with $v_i$ in $attrs$
                $recodings \leftarrow recodings \cup$ Anoymize($tuples_i,\ attrs',\ weight$)
            return $recodings$
for each $a_j$ of $attrs$
        $weight_j \leftarrow$ deep generalize for $a_j$
        $NCP_j \leftarrow$ Anonymize($tuples, attrs, weight_j$)
        $weightFinal \leftarrow$ Priority($NCP$)
return Anonymize($tuples, attrs, weightFinal$)
```

Figure 6. Optimized multidimensional K-Anonymity algorithm

IV. Experimental Results and Comparison

The experiments were performed using an Intel(R) Core(TM) i7-6700 CPU 3.40 GHz with 32 GB RAM. Adult data from the UCI Machine Learning Repository were used in the experiments. These data contain 1,000,000 tuples (expanded from 45,222 original tuples) and 14 attributes from U.S. Census data. We selected five attributes from this dataset as quasi-identifiers, as given in Table 1. The selected data have four categorical quasi-identifiers (i.e., Work class, Race, Sex, and Native Country) and one numerical quasi-identifier (i.e., Age).

TABLE 1. DESCRIPTION OF ADULT DATASET

Attribute	Distinct Values	Generalization
Age	74	MDR Tree
Work class	8	Hierarchy Tree
Race	5	Hierarchy Tree
Sex	2	Hierarchy Tree
Native Country	41	Hierarchy Tree

Considering that the numerical data would have outliers, which may increase the risk of re-identification, we preprocessed the raw data. For the numerical attribute, we found and merged the outlier of $\mu \pm 2\sigma$ into one group (i.e., a group of age less than 22 years and another of age more than 62 years).

In this experiment, we generated a hierarchy tree for the numerical attribute Age by using an MDR tree. The result is shown in Figure 7. This hierarchy tree has four levels, and the interval values of these four levels are *, one, two, and four. As mentioned in section 3, the numerical hierarchy tree is separated into four levels. By using the optimized multidimensional model proposed in this paper, the output data of the numerical attribute will follow this hierarchy tree. In other words, the numerical attribute of the output data will have four types of expression.

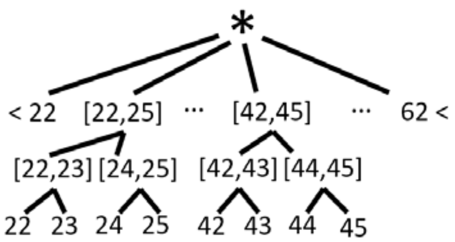


Figure 7. Experimental result of hierarchy tree of numerical attribute Age

Compared to the proposed model, the Mondrian algorithm uses a median partitioning method to segment the numerical attribute, and the output data would be unsuitable for analytical use because of the irregular intervals of the numerical attribute. Furthermore, owing to the regularity of hierarchical trees, future studies can consider merging the four-level hierarchy tree into a single-level hierarchy tree. In brief, depending on user demand, we can switch from four types of intervals to one fixed interval.

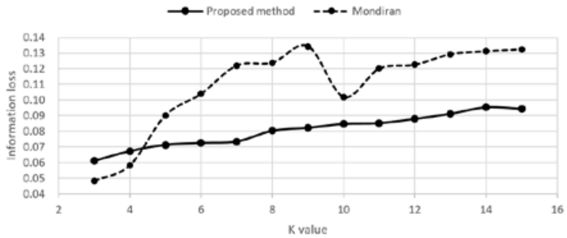


Figure 8. Comparison of information loss rate of proposed method and that of basic Mondrian algorithm

The present study enhances the previous studies' findings by providing a detailed development of the information loss metric and the hierarchy tree. Moreover, it compares the information loss rates of two models, namely, basic Mondrian algorithm and optimized multidimensional k-anonymity model. For a visual picture of the distinction, consider the graphical representation shown in Figure 8.

Intuitively, when the threshold of the k-value increases, the information loss rate increases in both cases. Furthermore, the proposed model showed significant gains over the basic Mondrian algorithm when the value of the threshold was more than 5, that is, the proposed model might outperform the basic Mondrian algorithm when the appetite for risk decreases (i.e., the k-value increases). However, direct comparisons must be treated with caution, so we suggest that the above statement be proved in future works.

V. Conclusions

The main conclusions of this study is can reduce the information loss rate effectively. Our research has suggested, albeit tentatively, that one potentially important influence of the priority weight-adjusted method on the privacy preservation process is a decrease in the information loss rate. The results have demonstrated that the proposed optimized multidimensional k-anonymity model can be implemented practically and can provide adequate results.

We must admit that this paper is still a preliminary study. In the future, we will strengthen the clustering of models through machine learning method, and enhance the testing method to measure the de-identification utility.

The issue of the characteristics of the "Best Privacy Protection Mechanism" is likely to puzzle researchers for some time to come [9], [20] - [22]. It is not within the scope of this paper to provide an extended discussion of the ongoing debates. Also, discussion of de-identifying the mixed attributes are beyond the framework of this paper.

References

[1] C. Dwork, "Differential Privacy," in *Automata, Languages and Programming*, 2006, pp. 1–12.

[2] C. C. Aggarwal and P. S. Yu, "A Condensation Approach to Privacy Preserving Data Mining," in *Advances in Database Technology - EDBT 2004*, 2004, pp. 183–199.

[3] H. Yang and I. Soboroff, "Privacy-Preserving IR 2015: When Information Retrieval Meets Privacy and Security," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA, 2015, pp. 1157–1158.

[4] T. Allard, D. Frey, G. Giakkoupis, and J. Lepiller, "Lightweight Privacy-Preserving Averaging for the Internet of Things," in *Proceedings of the 3rd Workshop on Middleware for Context-Aware Applications in the IoT*, New York, NY, USA, 2016, pp. 19–22.

[5] H. Yang, I. Soboroff, L. Xiong, C. L. A. Clarke, and S. L. Garfinkel, "Privacy-Preserving IR 2016: Differential Privacy, Search, and Social Media," in *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, NY, USA, 2016, pp. 1247–1248.

[6] A. Solanas and A. Ballesté, "V-MDAV: Variable group size multivariate microaggregation," *COMPSTAT'2006*, Jan. 2006.

[7] P. Samarati and L. Sweeney, "Protecting Privacy when Disclosing Information: k-Anonymity and Its Enforcement through Generalization and Suppression," 1998.

[8] L. Sweeney, "K-anonymity: A Model for Protecting Privacy," *Int J Uncertain Fuzziness Knowl-Based Syst*, vol. 10, no. 5, pp. 557–570, Oct. 2002.

[9] Z. Li, G. Zhan, and X. Ye, "Towards an Anti-inference (K, ℓ)-Anonymity Model with Value Association Rules," in *Database and Expert Systems Applications*, 2006, pp. 883–893.

[10] P. Samarati, "Protecting Respondents' Identities in Microdata Release," *IEEE Trans Knowl Data Eng*, vol. 13, no. 6, pp. 1010–1027, Nov. 2001.

[11] G. Aggarwal *et al.*, "Anonymizing Tables," in *Proceedings of the 10th International Conference on Database Theory*, Berlin, Heidelberg, 2005, pp. 246–258.

[12] L. Sweeney, "Achieving K-anonymity Privacy Protection Using Generalization and Suppression," *Int J Uncertain Fuzziness Knowl-Based Syst*, vol. 10, no. 5, pp. 571–588, Oct. 2002.

[13] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan, "Incognito: Efficient Full-domain K-anonymity," in *Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data*, New York, NY, USA, 2005, pp. 49–60.

[14] R. J. Bayardo and R. Agrawal, "Data Privacy Through Optimal k-Anonymization," in *Proceedings of the 21st International Conference on Data Engineering*, Washington, DC, USA, 2005, pp. 217–228.

[15] J. Xu, W. Wang, J. Pei, X. Wang, B. Shi, and A. W.-C. Fu, "Utility-based Anonymization for Privacy Preservation with Less Information Loss," *SIGKDD Explor Newsl*, vol. 8, no. 2, pp. 21–30, Dec. 2006.

[16] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan, "Mondrian Multidimensional K-Anonymity," in *Proceedings of the 22Nd International Conference on Data Engineering*, Washington, DC, USA, 2006, p. 25–.

[17] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan, "Workload-aware Anonymization," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2006, pp. 277–286.

[18] M.-A. Rizoiu, L. Xie, T. Caetano, and M. Cebrian, "Evolution of Privacy Loss in Wikipedia," *ArXiv151203523 Cs*, pp. 215–224, 2016.

[19] H. van Hoof, G. Neumann, and J. Peters, "Non-parametric Policy Search with Limited Information Loss," *J. Mach. Learn. Res.*, vol. 18, no. 73, pp. 1–46, 2017.

[20] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "L-diversity: Privacy Beyond K-anonymity," *ACM Trans Knowl Discov Data*, vol. 1, no. 1, Mar. 2007.

[21] H. Jian-min, Y. Hui-qun, Y. Juan, and C. Ting-ting, "A Complete (Alpha,K)-Anonymity Model for Sensitive Values Individuation Preservation," in *Proceedings of the 2008 International Symposium on Electronic Commerce and Security*, Washington, DC, USA, 2008, pp. 318–323.

[22] R. C.-W. Wong, J. Li, A. W.-C. Fu, and K. Wang, "(α, K)-anonymity: An Enhanced K-anonymity Model for Privacy Preserving Data Publishing," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2006, pp. 754–759.