THE INTERNATIONAL JOURNAL OF SCIENCE & TECHNOLEDGE

The K-Anonymity Approach for Privacy Preservation against Aggregate Knowledge Attacks

Poonam Joshi
Assistant Professor, Information Technology Department
Atharva College of Engineering, Mumbai, Maharashtra, India
Shubham Anand Wade
B.E. Information Technology Department
Atharva College of Engineering, Mumbai, Maharashtra, India
Chinmay Gajanan Wadkar
B.E. Information Technology Department
Atharva College of Engineering, Mumbai, Maharashtra, India
Deepukumar Ramji Yadav
B.E. Information Technology Department
Atharva College of Engineering, Mumbai, Maharashtra, India
Rohan Vishwas Salgaonkar
B.E. Information Technology Department
Atharva College of Engineering, Mumbai, Maharashtra, India

Abstract:
In todays world preserving the privacy of individuals is difficult as attacker may have abstract or aggregate knowledge about each record.This paper is about protecting privacy of individuals in publication scenarios using k-anonymity approach.The anonymization technique aims at generalizing attributes to ensure that aggregate values over the complete record will create equivalence classes of at size k.The anonymization technique addresses attack scenario and ensures that there is no information loss.

Keywords: *Privacy preservation,data mining,k anonymity*

1. Introduction

Now a days it is common that data published by organization or company is too detailed to expect the intruders to have accurate partial knowledge.Though an intruder might have some aggregate or abstract knowledge of a field.So we propose an anonymization technique that generalizes attributes.

2. Literature Survey

The basic methodhology is to form groups of records that have similar aggregate function values of their quasiidentifiers.
So we perform local generalizations independently within each group. It includes the following:
We define the problem of anonymizing data according to aggregate information.

- We define kf -*anonymity* that ensures privacy
- We define a utility-preserving anonymization algorithm;
- We verify our methods with real-world data and compare our results to Mondrian, a multidimensional localrecoding
- k-anonymity algorithm.

We group the records into equivalence classes, and perform *local-recoding generalization* on the values of each equivalence class independently.
Our method has two phases- In Phase one we divide the records into groups. We then form equivalence classes with respect to the function f. First, all records are sorted with respect to their f(qi2, qi2, ..., qin) value. Then, they are clustered into equivalence classes of sizes $k \leq |EC| \leq 2k - 1$.
To avoid overgenerelization of values we limit EC size.In the second phase we generalize values of each equivalence class separately.

2.1. Figures and Tables

Name	Income	Profit	Other income	Total Income
Michael	15	25	120	160
Bob	20	20	150	190
Adam	35	30	200	265
Tim	30	35	220	285

Table 1: Original Data

Id	Income	Profit	Other Income	Total Income
1	[15-20]	[20-25]	[120-150]	[150-200]
2	[15-20]	[20-25]	[120-150]	[150-200]
3	[30-35]	[30-35]	[200-220]	[250-300]
4	[30-35]	[30-35]	[200-220]	[250-300]

Table 2: Classical Anonymization Table

Id	Income	Profit	Other Income	Total Income
1	[15-20]	25	120	[160-190]
2	[15-20]	20	150	[160-190]
3	35	30	[200-220]	[265-285]
4	30	35	[200-220]	[265-285]

Table 3: Aggregate Anonymization Table

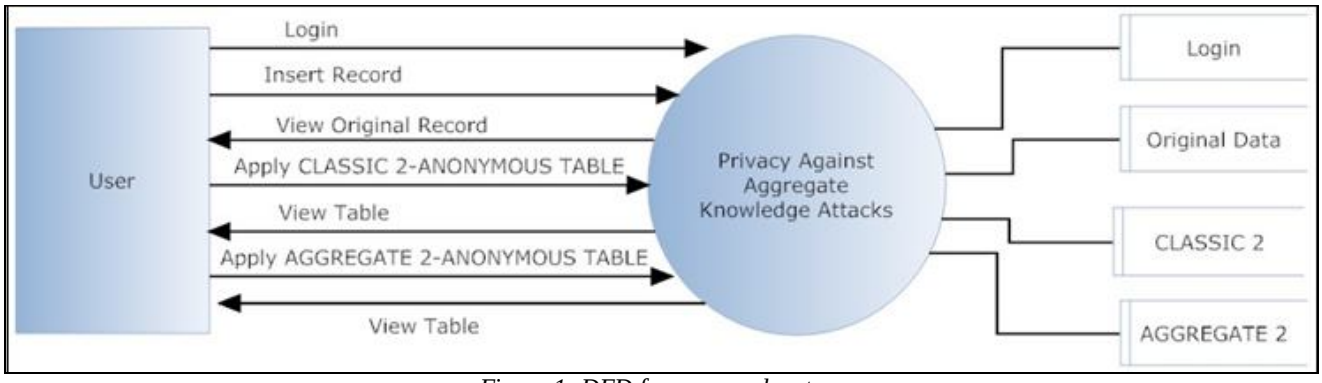

Figure 1: DFD for proposed system

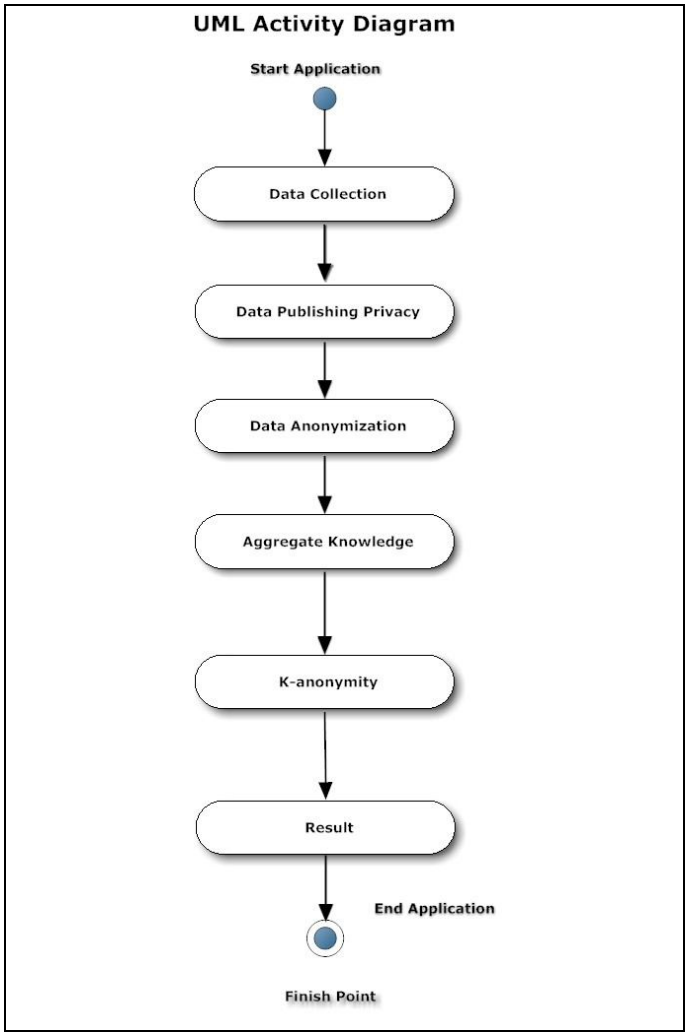

Figure 2: block diagram for proposed method

2.2. Equations

To estimate the loss of utility introduced by the value generalizations we use Normalized Certainty Penalty (NCP) metric . Let v be a value in original domain I. Then:

NCP(v) = { 0, v not generalized

|maxv − minv|/|I|, otherwise

where [minv, maxv] is the range to which v is generalized

2.3. Other Recommendations

Publishing data about individuals without revealing sensitive information about them is an important problem. In recent years, a new definition of privacy called *k*-anonymity has gained popularity. In a *k*-anonymized dataset, each record is indistinguishable from at least $k - 1$ other records with respect to certain identifying attributes. In this article, we show using two simple attacks that a *k*-anonymized dataset has some subtle but severe privacy problems. First, an attacker can discover the values of sensitive attributes when there is little diversity in those sensitive attributes. This is a known problem. Second, attackers often have background knowledge, and we show that *k*-anonymity does not guarantee privacy against attackers using background knowledge. We give a detailed analysis of these two attacks, and we propose a novel and powerful privacy criterion called ℓ-diversity that can defend against such attacks. In addition to building a formal foundation for ℓ-diversity, we show in an experimental evaluation that ℓ-diversity is practical and can be implemented efficiently.

3. Discussion & Conclusion

In Proposed system we aim at providing a form of k-anonymity to prevent attacks against identity disclosure. We propose a local-recoding generalization approach that preserves utility by generalizing the least number of values necessary to form equivalence classes of size k (or more) with respect to the aggregate function. Compared to classic k-anonymity, even **4.** for local-recoding methods, we achieve better utility as we do not create classes of completely identical records.

The proposed system forms a group of records that have similar aggregate function values of their quasiidentifiers. To achieve this we perform local generalizations independently within each group. We limit our discussion to numerical values, but our method can be extended to categorical if aggregate functions are defined over them.

- we propose a novel and powerful privacy criterion called ℓ-diversity that can defend against such attacks. In addition to building a formal foundation for ℓ-diversity, we show in an experimental evaluation that ℓ-diversity is practical and can be implemented efficiently.
- The proposed system forms a group of records that have similar aggregate function values of their quasiidentifiers. To achieve this we perform local generalizations independently within each group. We limit our discussion to numerical values, but our method can be extended to categorical if aggregate functions are defined over them.

In this paper we studied the problem of anonymizing data in the presence of aggregate knowledge. To address this attack we proposed a relaxation of k-anonymity, that we call kf - anonymity. We provided a utility-preserving algorithm which greedily selects a solution that satisfies our guarantee.

To the best of our knowledge, this is the first work treating aggregate information as potential attacker knowledge. In the future, we will extend our guarantee to provide a form of l-diversity. We also wish to examine more complicated functions as potential background knowledge and also attack scenarios where the attacker has knowledge of multiple aggregate values for a record. Finally, we will examine the scenario that an attacker has partial knowledge of a record, i.e., some attribute values, additionally to her aggregate-knowledge.

4. References

1. L. Sweeney, "k-Anonymity: A Model for Protecting Privacy," IJUFKS, vol. 10, no. 5, 2002.
2. Meyerson and R. Williams, "On the Complexity of Optimal Kanonymity," in PODS, 2004, pp. 223–228.
3. J. Xu, W. Wang, J. Pei, X. Wang, B. Shi, and A. Fu, "Utility-Based Anonymization Using Local Recoding," in KDD, 2006, pp. 785–790.
4. "Uci repository," http://archive.ics.uci.edu/ml/datasets.html.
5. K. LeFevre, D. J. DeWitt, and R. Ramakrishnan, "Mondrian Multidimensional k-Anonymity," in ICDE, 2006.
6. "Utd anonymization toolbox," http://cs.utdallas.edu/dspl/cgibin/ toolbox/.
7. P. Samarati and L. Sweeney, "Generalizing Data to Provide Anonymity when Disclosing Information (abstract)," in PODS (see also Technical Report SRI-CSL-98-04), 1998.
8. G. Aggarwal, T. Feder, K. Kenthapadi, R. Motwani, R. Panigrahy, D. Thomas, and A. Zhu, "Approximation Algorithms for k-Anonymity,"Journal of Privacy Technology, 2005.
9. H. Park and K. Shim, "Approximate algorithms for k-anonymity," in SIGMOD, 2007, pp. 67–78.
10. K. LeFevre, D. J. DeWitt, and R. Ramakrishnan, "Incognito: Efficient Full-domain k-Anonymity," in SIGMOD, 2005, pp. 49–60.
11. R. J. Bayardo and R. Agrawal, "Data Privacy through Optimal k- Anonymization," in ICDE, 2005, pp. 217–228