A State Decision Tree based Backtracking Algorithm for Multi-Sensitive Attribute Privacy Preserving

Yanchao Zhang, Institute of ICT Security Research, China Academy of Information and Communications Technology, Beijing, China
Qing Liu, Institute of ICT Security Research, China Academy of Information and Communications Technology, Beijing, China
JunJun Cheng, China Information Technology Security Evaluation Center, Beijing, China
JiJia Yang, Institute of ICT Security Research, China Academy of Information and Communications Technology, Beijing, China

ABSTRACT

Beyond l-diversity model, an algorithm (l-BDT) based on state decision tree is proposed in this paper, which aims at protecting multi-sensitive attributes from being attacked. The algorithm considers the whole situations in equivalence partitioning for the first, prunes the decision tree according to some conditions for the second, and screens out the method with the least information loss of equivalence partitioning for the last. The analysis and experiments show that the l-BDT algorithm has the best performance in controlling the information loss. It can be ensured that the published data is the most closed towards the original data, so as to ensure that the published data is as useful as possible.

KEYWORDS

1-diversity, Data Security, Multi-sensitive, Privacy Protection

1. INTRODUCTION

With the increasing demand of the sharing service, how to protect the users’ privacy become the focus of the all concerned parties. In the current researches, common approach to this problem is to generalize the quasi identity attribute (such as k-anonymity), or to vary the sensitive attribute of an equivalence class (such as 1-diversity) to lower the risk of exposure of any private data. Until now, 1-diversity and its improved model has been proved to be effective, and become the mainstream of research in such field. Its works by keep any sensitive attribute in one equivalence class appears at a frequency of no higher than 1/l. In this way even if the attacker fixates the target’s equivalence class, the highest possibility of his deduction of the exact record will also be just 1/l, hence increasing the difficulty of deducting a certain sensitive attribute on the basis of quasi identity attribute and background knowledge. However in order to make the data to be posted meet the standards of the 1-diversity model, some data will be left out in the links of quasi identity attribute and the sensitive attribute, which is unavoidable. This will undermine the integrity of the original data. For instance, a hospital gives publicity of the information of some patients for the research purpose to some concerned institutions. For the sake of the privacy of such patients, the hospital would adopt the 1-diversity Algorithm. During the process some data will possibly be dropped by the model because of they do
not pass the Algorithm. Because of the data is not complete, the institution will not be able to conduct an accurate research on the mobility of the diseases. Therefore, for the protection of the sensitive information, many scholars have proposed abundant models and Algorithms. Yang Xiaochun is one of them. Her idea is based on the thought of multi-dimensional barrel and she further put up three Algorithms on the basis of greed. In this article, an algorithm to protect varied sensitive data according to the rank of the sensitivity of the data. The Algorithm gives different coefficients to different data to protect the privacy respectively. A model of (K, 1) was mentioned. It would decide the anonymity and the sensitivity of every tuple and achieve distinction of individual anonymity and sensitivity. In this article, a t-closeness has been put forward to make up the shortcoming of the 1-diversity model after actually application of the K-anonymity and 1-diversity. This article [9-13] is about an improved Algorithm of the 1-diversity model. But currently the hidden rate of the data is fairly high, especially when there are multiple sensitive attributes. When too much data is dropped in the process, the integrity of the data is damaged, which unavoidably cause the inadmissibility of the data.

The 1-BDT Algorithm, while meeting the requirement of the 1-diversity model, it will lower the data to be dropped during the calculating process to the minimum, so that the integrity of the data will be maintained. 1-BDT Algorithm will organize all the possible divisions of an equivalence class by quadratic tree structure, and trace back along and trim the tree “branches” to rule out the divisions that don’t achieve minimum data losses. Meanwhile, the Algorithm will keep updating the best solution and achieve the minimum loss of data during its process.

2. DEFINITION AND ANALYSIS OF CONCERNED PROBLEMS

Let us assume that a user will publish a relation T\{A1,A2,……Ap,S1,S2,……,Sd\}.Ai(1≤i≤n) means quasi identity attribute, Sj(1≤j≤d), sensitive attribute. Let’s suppose that there are n records in T, meaning |T|=n, and every record will be t_i(1≤j≤n). And t[X] indicating the X attribute of t.

Definition 1(individual sensitive attribute l-diversity group[4]): record a group of sensitive attribute as G, suppose v is the biggest value of frequency in the group G, c(v) is the value of the frequency, if \[\frac{c(v)}{|G|} \leq \frac{1}{l}\] (|G| is the number of records in G), then G applies to l-diversity.

Definition 2(multiple sensitive attribute l-diversity group[4]): groups of T are GT{G1,G2,……,Gm}, say every group is Gi(1≤i≤m) which all apply to l-diversity, the GT is the group on T that applies to l-diversity.

Theorem 1: the groups applying to l-diversity will at least have l elements.

Proof: suppose there are n elements in a group, and n<l, then n elements certainly cannot ensure that any l elements applying to l-diversity. As a result, the group applying to l-diversity should at least have elements.

Definition 3 (tuple satisfying l-diversity): a group having l elements and applying to l-diversity, then this group is the tuple satisfying l-diversity, named tuple.

Theorem 2: a group g satisfying l-diversity, when a random element eliminated from it, the group we get is g’ whose elements are still more than l, then the group g’ satisfies l-diversity.

Proof: suppose there are n elements in the group g, namely e1, e2……en. Group g satisfies l-diversity. Then the values of any 2 elements out of e1, e2……en are different at any dimension. When a random element is taken out of the group, any random two elements out of the left n-1 elements still can satisfy that. Since the number of the elements in the group g’ are still no less than l, the group g’ satisfies l-diversity.

Theorem 3: a group satisfying l-diversity. If this group is divided into n sub-groups, the number of whose elements are all no less than l, then the n sub-groups all satisfy l-diversity.

Proof: group g satisfies l-diversity. When it is divided into g1, g2…gm, any random sub-group gi(1≤i≤m) has more than one element. In this way, according to any gi, a deduction that group g satisfies l-diversity can be made. In addition to the elements contained in the packet G1, other
Simulation of Competition in NGNs with a Game Theory Model
[www.igi-global.com/chapter/simulation-of-competition-in-ngns-with-a-game-theory-model/108096?camid=4v1a](www.igi-global.com/chapter/simulation-of-competition-in-ngns-with-a-game-theory-model/108096?camid=4v1a)

The Diffusion of WiMax Technology: Hurdles and Opportunities
[www.igi-global.com/chapter/diffusion-wimax-technology/21690?camid=4v1a](www.igi-global.com/chapter/diffusion-wimax-technology/21690?camid=4v1a)