基于 K-匿名的快递信息隐私保护应用*

韦 茜, 李星毅
(江苏大学 计算机科学与通信工程学院, 江苏 镇江 212013)

摘 要: 针对快递单号被盗取和快递单信息保护不当造成的隐私泄露问题进行了研究,提出了一种新型 K-匿名模型对快递信息进行匿名处理。该方法通过随机打破记录中属性值之间的关系来匿名数据,相比于其他传统方法,克服了数据间统计关系丢失的问题和先验知识攻击。实验结果表明,新型 K-匿名方法能够加强隐私保护和提高知识保护的准确性。

关键词: 快递单; K-匿名模型; 隐私保护; 数据匿名
中图分类号: TP309.2 **文献标志码**: A **文章编号**: 1001-3695(2014)02-0555-03
doi:10.3969/j.issn.1001-3695.2014.02.056

Express information protection application based on K-anonymity

WEI Qian, LI Xing-yi
(*School of Computer Science & Telecommunication Engineering*, *Jiangsu University*, *Zhenjiang Jiangsu* 212013, *China*)

Abstract: Courier tracking number stolen and express list information are not protected carefully that may cause privacy leakage. So this paper proposed a new K-anonymity model to protect express information. Unlike most anonymity methods, this method anonymized data by randomly breaking links among attribute values in records. By data randomization, this method maintained statistical relations among data to preserve knowledge, whereas in most anonymization methods, knowledge was lost. Further more, it proposed the algorithm for extra privacy protection to tackle the situation where the user's prior knowledge of original data may cause privacy leakage. Simulation results show that the new K-anonymity method can enhance privacy protection and improve the accuracy of the knowledge protection.

Key words: express list; K-anonymity model; privacy protection; data anonymization

随着网购的活跃,快递业空前发展,由于我国快递行业的各项法律政策以及各快递公司对客户的隐私安全保护并非十分完善,因此消费者的隐私安全受到了威胁。快递行业中造成消费者隐私泄露的原因有很多,本文以快递单信息为研究对象讨论快递业存在的隐私泄露问题。在快递包裹被运送的整个流程中,快递单号是快递包裹的唯一标志代码,由数字和字母组成,用于快递公司、发件人以及收件人跟踪快递信息。快递单号在整个递送流程中始终是公开的,在相关网站输入快递单号,就能查询到快递公司名称、收发货地点、家庭住址及客户手机号等,造成客户隐私泄露。另外消费者收到货物后,将贴有快递单的包装盒随意丢弃,也易造成个人隐私泄露。如何保护快递单中的客户信息安全是亟待解决的问题。目前隐私保护技术在数据库的应用主要集中在数据挖掘和匿名发布两大领域[1]。本文主要讨论数据匿名化的应用。现在国内关于隐私保护技术的研究主要集中于基于数据失真或数据加密技术方面,如基于隐私保护分类挖掘算法、分布式数据的隐私保持协同过滤推荐等。研究工作者一直致力于寻找更好实现隐私保护原则和算法之间平衡的方法。总之,国内关于隐私保护的技术研究还处在初步阶段,具有广阔的研究发展空间。

对快递单中消费者信息进行数据匿名化时,需要将家庭住址和电话信息保留,否则货物无法准确送达。但引发另一问题,若有人提前看过原始数据,即具有先验知识,纵使后来将数据匿名,仍有隐私泄露的可能。因此保护用户的隐私和抵御先验知识攻击,是本文研究的主要内容。本文针对快递业的特殊性,利用新型 K-匿名隐私保护模型解决快递中心保护客户隐私安全的问题。

1 相关工作

1.1 快递中的业务流程

图 1 为简略的快递过程。其中造成客户隐私泄露的环节包括两个方面:a)从快递公司本身来说,快递公司将贴有快递单的货物分发到快递公司各地配送中心的过程中,工作人员都可以通过快递单号在内网查询到客户的详细信息,此环节中若没有严格的权限限制,易造成客户隐私泄露;b)在各中心快递员将货物派送到收货人的环节中,由于快递员能第一手获得客户信息,而各快递公司所聘快递员素质良莠不齐,也易造成客户信息的泄露。因此无论是哪一环节中的工作人员获得消费者的信息越少越好。

本文将主要解决第一个环节可能带来的隐私泄露,即作为快递中心本身来说,如何保护客户的隐私安全。

1.2 相关概念

K-匿名模型是一种简单有效的隐私保护模型,通常采用泛化和隐匿技术[2]来实现该模型,但该模型不能抵制同质性攻

收稿日期: 2013-03-22; **修回日期**: 2013-05-17 **基金项目**: 国家"十一五"科技支撑计划资助项目(2006BAG01A0);国家自然科学基金资助项目(10972027);江苏大学校基金资助项目(11JDG064)

作者简介:韦茜(1989-),女,江苏扬州人,硕士研究生,主要研究方向为隐私安全保护与图像处理(916884324@qq.com);李星毅(1969-),男,江苏镇江人,副教授,硕导,主要研究方向为智能交通、复杂系统智能分析.

击和背景知识攻击[3]。Han 等人[4]提出了 L-多样性算法,它保证每个等价类中敏感信息足够多样,以抵制同质性攻击和背景知识攻击,但该模型不能抵制偏斜攻击和相似性攻击。Truta 等人[5]提出 p-Sensitive k-anonymity 模型,它简化了 L-多样性实现中的参数设置,但该模型没有控制等价类中敏感值出现的频率,当 k 比 p 大很多时,无法抵制概率攻击。此后,Wong 等人[6]提出了(α,k)-匿名模型,要求每个等价类的敏感值频率不大于 α,但是(α,k)-匿名模型为所有的敏感值设置统一的频率约束,适应性比较差,因为数据表中不同的敏感值可能需要不同的频率约束。Li 等人[7]提出了 T-closeness 框架,方法要求每个等价类的敏感值的分布要接近于其在原始数据表中的分布。但它的缺点也很明显,即使满足与原数据分布相似也并不一定能保证发布数据的安全性,而且它将很大程度地破坏数据的实用性,一个完美的 T-closeness 将完全割裂敏感属性和准标志符之间的关联。

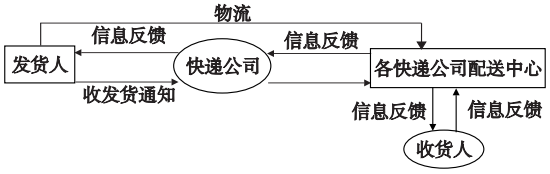

图1 快递业务流程

1.3 K-匿名模式

文中对快递单的隐私保护主要是针对客户而言,所以发件人的姓名、邮编,以及货物等其他信息暂不作考虑,收件人家庭详细住址不具体列出。表 1 和 2 列出了常见快递单上包含的一些基本信息,表 1 中的快递单号以申通为例,申通单号一般由 12 位数字组成,目前常见以 268 * 、368 * 、468 * 等开头。其中收件人手机号、住址都设定为敏感信息即多敏感属性[8,9]。

表 1 快递单收件人信息

收件人姓名	QI	敏感数据	
	快递单号	收件人手机	收件地址
张雨	268853074869	15205262911	浙江温州
李清	368853054869	15205262911	江苏扬州
李宁	268854474869	18005262911	河南洛阳
张红	368853558691	18096052629	江苏苏州
李明	368993074869	18796052629	山西太原

表 2 快递单寄件人信息

QI		
寄件人姓名	收件人手机	收件地址
王铭	15205262911	浙江温州
李玉红	15205262911	江苏扬州
赵波	18005262911	河南洛阳
张一	18096052629	江苏苏州
孙云玲	18796052629	山西太原

表 3 中的 * 代表任意信息。由于 K-匿名模式的简洁性和适用于多种算法的属性,使它在数据发布中变得非常流行。但是,在受到攻击的情况下,K-匿名模式仍然会泄露一些敏感信息,因此,它并不能完全保证隐私的安全性。

在对快递单上的信息匿名之前,先假设快递公司制表员打印出来的快递单子上只有收件人电话和收件人地址,虽然快递单号是包裹的唯一标志,为了安全起见还是将它隐匿,因为在快递单的右上角有一条形码,用巴枪或 DIAD(速递资料收集器)进行扫描就可以获得包裹的详细信息。对于快递公司来说,快递单号要留存,在货物被揽收之后,可以以手机短信的方式将单号发送到客户的手机上,以便客户跟踪物流信息。一旦客户签收后,交易将关闭,若再以快递单号到快递公司官网上查询时,显示的信息将是已匿名的。此后任何人包括快递公司的工作人员利用快递单号查询货物信息时,显示的都将是已经匿名的信息,这样网上存在的快递单号不法交易也将无处藏身。今后基于信息共享、科学研究等方面的需要,若直接将消费者的数据进行发布,会造成用户敏感信息的泄露。本文采用的新型 K-匿名方法,既可以让快递中心做到保护消费者的隐私安全,也可以在数据共享时,降低消费者隐私信息泄露的风险。

表 3 对快递单的一般 K-匿名

QI			敏感数据	
快递单号	寄件地址	寄件人手机	收件人手机	收件地址
268 *	江苏 *	187 *	152 *	浙江 *
368 *	湖北 *	187 *	152 *	江苏 *
268 *	江苏 *	187 *	180 *	河南 *
368 *	浙江 *	152 *	180 *	江苏 *
368 *	湖南 *	152 *	187 *	山西 *

2 基于 K-匿名的新型隐私保护方法

由表 3 可得出,虽然 K-匿名一定程度上减小了用户与它们的敏感数据的关联,若用户具有 QI(准标志符)和敏感信息关联的先验知识,就会加剧敏感属性的泄露,而一般的 K-匿名无法解决这个问题。下面介绍的这种新型的 K-匿名保护模型,不同于 K-匿名模型和 L-diversity 方法,它打破 QI 与敏感属性之间的关联,随机对数据集进行匿名,匿名原则是用原表中的数据替代将要被匿名的对象,这在一定程度上保证了数据的实用性。当用 SQL 查询 RA 算法匿名的数据集时,相对误差会小很多。

定义 1 给定一个数据集 D,它的准标志符为 QI,Q-I 为所有记录中 QI 的集合。

定义 2 以表 1 为例,算法中的属性组 Q-I 的属性 $Q=\{Q_1,Q_2,Q_3\}=\{$收件地址,收件手机号,快递单号$\}$。

属性 Q_1 各值所占比例如表 4 所示。

表 4 属性 Q_1 各值所占比例

Q_1 的值	浙江	江苏	河南	山西
Dist_1	1/5	2/5	1/5	1/5

Q_2、Dist_2、Q_3、Dist_3 的值依此类推。

定义 3 属性组合 q 中包含的值在数据集 D 中占的比例被表示成 $\sup p_D(q)$。

定义 4 相对差异即 $R(q)=\dfrac{(\sup p_{D'}(q)-\sup p_D(q))}{\sup p_D(q)}$,相对偏差 $B(q)=|E(R(q))|$,这为 $\sup p_D(q)$ 和 $\sup p_{D'}(q)$ 之间的差异提供了一个测量标准。并且 $B(q)$ 越小,知识发现越准确。

相对差的方差:

$$\mathrm{var}(R(q))=\frac{1}{m^2}\sum_{i=1}^{|q|}\left(\frac{\sup p_D(q\Theta q_i)^2}{\sup p_D(q)^2}(P(Q_i=q_i)-P(Q_i=q_i)^2)\right) \tag{1}$$

由式(1)可以看出,$\mathrm{var}(R(q))$ 与 $|q|$ 之间呈线性关系,随着 $|q|$ 的增大而增大。也就是说,$\sup p_D(q)$ 与 $\sup p_{D'}(q)$ 之间的相对差随着属性值组合的增长而变得更加不确定。属性值组合的长度越小,准标志符与敏感数据被保护得越好,知识发

现的准确率越高[10]。

基于 K-匿名的新模型算法:

```
RA 算法
输入:设原始数据集为 D, Q 是 Q-I 组的属性,λ(λ≤m)是将被匿名属性的个数。
输出:D 表示被匿名后的数据覆盖。
begin
m:=|Q|;
n:=|D|;
Dist_i:=Φ;
for i:=1 to m  do
begin
  Dist_i:= Qi 的值;
end
for j:=1 to n do
begin
  在第 j 个记录中,以相同的概率随机从 Q-I 属性中选择 λ
  for k:=1 to λ do
  begin
    attr:=  Q-I 的 k 个属性被选择;
    基于 Dist_i 随机生成一个值赋给 attr;
    用一个新值代替 attr 的值;
    end
  end
end
```

算法在 λ 增大时能够增强隐私保护,当 λ 增大时,$R(q)$ 和 $\mathrm{var}(R(q))$ 也随之增大。本文在 Q-I 中随机选取将要被匿名的属性 λ 的概率为 $1/\binom{m}{\lambda}$。在上述算法中,一个属性值组合不能被匿名成其他形式。如果两个属性值组合 q 与 q' 有 η 个不同属性,并且这两个属性值组合能够相互转换,那么被匿名的应该是这 η 个不同的属性。下面将讨论一下 λ 对 $P(q\rightarrow q')$ 与 $P(q'\rightarrow q)$ 的影响。用新算法对快递表单的匿名如表 5 所示。

表 5　用新算法对快递表单的匿名

QI			敏感数据	
快递单号	寄件地址	寄件人手机	收件人手机	收件地址
268853074869	江苏镇江	18786232459	18005262911	浙江温州
368853054869	浙江温州	18796020628	15205262911	江苏扬州
268854474869	江苏南京	18796052629	15205262911	江苏苏州
368853558691	湖北武汉	15268952036	18096052629	河南洛阳
368993074869	湖南邵阳	15205876216	18796052629	山西太原

由表 5 可以得出,新模型将准标志符和敏感数据之间的联系随机打破,并隐匿标志符收件人姓名,没有像传统模型将标志符与准标志符都隐匿,使数据丢失并失去实用性(被隐匿的数据已用黑体标出)。

定理 1　假设 λ 是每条记录中将要被匿名属性的个数。令 $q\in D, q'\in D'$,q 与 q' 是 QI 的子集 $|q|=|q'|\leqslant m$,它们两个能互相转换,$P(q\rightarrow q')$ 与 $P(q'\rightarrow q)$ 的概率随着 λ 的增大而增大。

证明　假设 q 与 q' 相差 η 个属性,$\eta\leqslant\lambda$。令 $\mathrm{AttrSet}^{\eta+k}$ 为 $\eta+k$ 个属性所有可能的组合,其中 $\mathrm{AttrSet}^{\eta+k}$,$i=1\cdots$,$\binom{|q|-\eta}{k}$,$k\leqslant|q|-\eta$,$q$ 为 QI 的任意属性集。令 $P(\mathrm{AttrSet}^{\eta+k}, q')$ 为 $\eta+k$ 个属性被匿名为相对应属性集 q' 的概率,则 $P(q\rightarrow q')=\sum_{k=0}^{\min(\lambda-\eta,|q|-\eta)}\left[\frac{\binom{m-|q|}{\lambda-\eta-k}}{\binom{m}{\lambda}}\right]\sum_{i=1}^{\mathrm{AttrSets}^{\eta+k}}P(\mathrm{AttrSets}_i^{\eta+k},q')=\sum_{k=0}^{\min(\lambda-\eta,|q|-\eta)}\mathrm{Pr}_k^{\lambda,\eta}$。其中 $\mathrm{Pr}_k^{\lambda,\eta}$ 表示 $\left[\frac{\binom{m-|q|}{\lambda-\eta-k}}{\binom{m}{\lambda}}\right]\sum_{i=1}^{\mathrm{AttrSets}^{\eta+k}}P(\mathrm{AttrSets}_i^{\eta+k},q')$,当 λ 增加 1 时,得到

$$\frac{\mathrm{Pr}_k^{\lambda+1,\eta}}{\mathrm{Pr}_k^{\lambda,\eta}}=\frac{\binom{m-|q|}{\lambda+1-\eta-k}/\binom{m}{\lambda+1}}{\binom{m-|q|}{\lambda-\eta-k}/\binom{m}{\lambda}}=\frac{(\lambda+1)(m-\lambda+1)}{(\lambda+1-\eta-k)(m-\lambda+1-|q|+\eta+k)}$$

其中:$1\leqslant\frac{\mathrm{Pr}_k^{\lambda+1,\eta}}{\mathrm{Pr}_k^{\lambda,\eta}}\leqslant\lambda+1$,且 $\eta+k\leqslant\min(\lambda,|q|)$。很明显 $P(q\rightarrow q')$ 随着 λ 而增减。$P(q'\rightarrow q)$ 的证明类似。

$$B(q')=\left|\sum\left[\frac{\sup p_D(q)}{\sup p_D(q')}P(q\rightarrow q')\right]-P(q'\rightarrow q)\right| \tag{2}$$

$$\mathrm{var}(R(q'))=\sum\left(\frac{\sup p_D(q)}{\sup p_D(q')}\right)^2(P(q\rightarrow q')-P(q\rightarrow q')^2)+(P(q'\rightarrow\neg q')-P(q'\rightarrow\neg q')^2)] \tag{3}$$

由定理 1 和式(2)(3)可以得出,通过改变 λ 的值还可以来调节平均偏差和方差。也就意味着知识发现的准确性也和 λ 有关,这一点对于本文算法的有效性十分重要。

3　实验与分析

3.1　实验数据选取及实验步骤

本实验采用申通快递客户资料的数据集,总共包括 3 162 条记录。本文先用 RA 和 L-diversity 算法对数据集进行匿名,再对匿名过的两个数据集随机进行 SQL 查询,并比较两者的结果平均准确率。

对于本文的 RA 算法,查询形式类似于[11]:

SELECT count(*) FROM tablename

WHERE Q_1 in R_1 AND $\cdots$ Q_m in (R_m) AND S in R_s

其中:Q_i 是 QI 的第 i 个属性,S 是敏感属性,R_i 是第 i 个属性的查询结果。

但是用 L-diversity 算法进行匿名得到的数据集是广义的,所以不能直接采用上述查询模式,因而每条记录的查询使用下列方式获得:

$$\sum_p\left(\#_p(R_s)\prod_{i=1}^{m}\frac{\#_p(R_i\cap\mathrm{values}(Q_i))}{\#_p(\mathrm{values}(Q_i))}\right) \tag{4}$$

通过式(4)可以得到一个估计后的查询结果。

本文总共用 10 个数据集来做实验,对于第 $j(1\leqslant j\leqslant 10)$ 个数据集,对它用两种方法匿名后,从中随机选取敏感属性和 j 个属性的组合实施 100 次查询。对于每个选定的属性,本文选择分类值的一个随机数的域作为 R_i。实验中的每个数据集的查询结果都会计算出一个相对误差为 $\frac{|\mathrm{Ans}(\mathrm{query},D)-\mathrm{Ans}(\mathrm{query},D')|}{\mathrm{Ans}(\mathrm{query},D)}$,其中 $\mathrm{Ans}(\mathrm{query},D)$ 为对数据集 D 返回的查询结果。

3.2　实验环境

硬件环境:PC 主频 1 GHz,处理器 AMD Athlon™ Ⅱ x255,内存 2 GB。操作系统:Microsoft Windows XP。编程环境:Eclipse + Oracle 10g 数据库。

3.3　实验结果对比分析

由图 2 可以看出,当查询的属性更敏感更多时,用 RA 算法匿名的数据产生的相对误差比 L-diversity 小很多。这是因为在假定每个 QI 属性均匀分布的基础上,用式(4)对 L-diversity 算法匿名的数据集的查询结果进行预测。随着高频率与低频率出现的值的组合比例减小,而包含的属性越多时,相对误差上升的速率会减慢。

图 3 要说明的是,在 RA 算法中属性关联被泄露的风险是受 λ 值影响的。随着 λ 的增大,原始数据之间的关联会越来越小,数据之间的虚假关联越来越大。(下转第 567 页)

参与者 A、B 最终达到期望纳什均衡点:

$$(\frac{1}{2}(u_A^+ - u_A^-) - \frac{1}{2}F_{A_1}, \frac{1}{2}(u_A^+ - u_A^-) - \frac{1}{2}(F_B + F_{A_2}))$$

多方情形的证明可在此基础上进行推广。所以根据理性公平性的定义,此时博弈协议满足理性公平性。

3.4　鲁棒性

理性参与者在不完美信息博弈中,总是为了最大化自己的利益进行策略选择。下面对鲁棒性[10,11]反应协议在异常中断情况下的健壮性进行证明。

定理 4　即使基于信息熵的理性交换协议异常中断,也能保证参与者的公平性,即本文方案具有鲁棒性。

在协议进行中的第一轮,若 A 选择 quit_A,则 B 不做任何动作。A 和 B 的收益均为 0,是公平的。

在协议的第二轮,若 B 在收到 A 发送的消息后选择 quit_B,此时 B 尽管将收到的消息 m_1 通过验证解密过程得到 $E_{k_A}(\text{item}_A)$,没有密钥 k_A 所以无法得到 item_A。这种情况下 B 收益为 0,同时对 A 不会造成损失,满足公平性。

在协议的第三轮,若 A 在收到 B 发送的消息后选择 quit_A,B 经过一段时间后仍没收到 m_3 则提起申诉,使 A 收到惩罚值 F_A,此时 A 的期望收益是 $\frac{1}{2}(-F_{A_1} - 2F_{A_2} + u_A^+ - u_A^-)$,显然小于发送 m_3 时的期望收益 $\frac{1}{2}(u_A^+ - u_A^-) - \frac{1}{2}F_{A_1}$。理性参与者 A 不会选择这样的策略,否则与假设矛盾。所以第三轮,A 会发送消息 m_3,最终与双方的期望收益保持一致,满足公平性。

综上,该协议在异常中断时,仍能保证公平性,所以本文的方案具有鲁棒性。

4　结束语

本文将极大熵原理引入理性交换协议中,构建了一个完全不完美信息动态博弈协议模型,结合期望收益函数和期望均衡,重新给出理性交换协议的公平性描述;在此基础上基于信息熵设计了一个理性交换协议,对协议安全性和公平性进行了证明与分析。结果表明该协议能达到期望均衡,提高效率的同时实现理性公平性,具有很好的适应性。

参考文献:

[1] ASOKAN N. Fairness in electronic commerce[D]. Waterloo: University of Waterloo, 1998.

[2] SYVERSON P. Weakly secret bit commitment: applications to lotteries and fair exchange[C]//Proc of the 11th IEEE Computer Security Foundations Workshop. 1998: 2-13.

[3] BUTTYÁN L, HUBAUX J P, ČAPKUN S. A formal model of rational exchange and its application to the analysis of Syverson's protocol [J]. Journal of Computer Security, 2004, 12(3-4): 551-587.

[4] ALCAIDE A. Rational exchange protocols[D]. Leganés: Universidad Carlos III de Madrid, 2008.

[5] ALCAIDE A, ESTEVEZ-TAPIADOR J M, HERNANDEZ-CASTRO J C, *et al*. Cryptanalysis of Syverson's rational exchange protocol[J]. International Journal of Network Security, 2008, 7(2): 151-156.

[6] 姜殿玉. 带熵博弈论及其应用[M]. 北京: 科学出版社, 2008.

[7] ASHAROV G, CANETTI R, HAZAY C. Towards a game theoretic view of secure computation[C]//Proc of the 30th Annual International Conference on Theory and Applications of Cryptographic Techniques: Advances in Cryptology. Berlin: Springer-Verlag, 2011: 426-445.

[8] LI Guo-qiang, GU Yong-gen, TAO Xiu-ting, *et al*. A game theoretic model and tree analysis method for fair exchange protocols[C]//Proc of the 5th International Symposium on Theoretical Aspects of Software Engineering. 2011: 243-246.

[9] BURROWS M, ABADI M, NEEDHAM R M. A logic of authentication[J]. ACM Trans on Computer Systems, 1989, 8(1): 18-36.

[10] ASOKAN N, SHOUP V, WAIDNER M. Optimistic fair exchange of digital signatures[J]. IEEE Journal on Selected Areas in Communications, 2000, 18(4): 593-610.

[11] 彭长根. 面向群体的签名、签密和签约的研究[D]. 贵阳: 贵州大学, 2007.

(上接第 557 页)

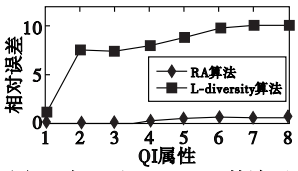

图2　在RA和L-diversity算法下数据查询结果的相对误差比较

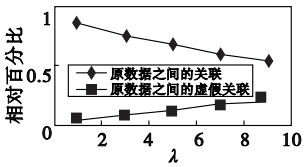

图3　原数据之间的关联程度与λ之间的关系

4　结束语

隐私信息安全是快递业中面临的重要问题。传统隐私研究技术在抵御先验知识攻击和保证数据实用性方面并没有特别好的解决方案。针对这两点,文中给出了基于 K-匿名的新型隐私模型,并给出了具体实现算法 RA。在实际数据集上进行了实验,结果表明新算法在隐私保护水平和数据实用性方面较传统 L-diversity、K-匿名算法有明显改善。此外若将 RA 算法集成其他匿名技术可能实现隐私保护与知识保护之间的平衡,这有待进一步研究。

参考文献:

[1] 周水庚,李丰,陶宇飞,等. 面向数据库应用的隐私保护研究综述[J]. 计算机学报, 2009, 32(5): 847-860.

[2] 魏琼. 数据发布中的隐私保护方法研究[D]. 武汉: 华中科技大学, 2008.

[3] 韩建明,于娟,虞慧群,贾洄. 面向敏感值的个性化隐私保护[J]. 电子学报, 2010, 38(7): 1723-1728.

[4] HAN Jian-min, YU Hui-qun, YU Juan. An improved L-diversity model for numerical sensitive attributes [C]//Proc of the 3rd International Conference on Communications and Networking. 2008: 938-943.

[5] TRUTA T M, UINAY B. Privacy protection: *p*-sensitive *k*-anonymity property [C]//Proc of the 22nd International Conference on Data Engineering Workshops. New York: ACM Press, 2006: 94.

[6] WONG R C W, LI Jiu-yong, FU A W C, *et al*. (α, k)-anonymous data publishing [J]. Journal of Intelligent Information Systems, 2009, 33(2): 209-234.

[7] LI Ning-hui, LI Tian-cheng, VENKATASUBRAMANIAN S. T-closeness: privacy beyond K-anonymity and L-diversity[C]//Proc of the 23rd International Conference on Data Engineering. [S. l.]: IEEE Press, 2007: 44-56.

[8] 朱青,赵桐,王珊. 面向查询服务的数据隐私保护算法[J]. 计算机学报, 2010, 33(8): 1315-1323.

[9] 刘腾腾,倪巍伟,崇志宏,等. 多维数值敏感属性隐私保护数据发布方法[J]. 东南大学学报, 2010, 40(4): 699-703.

[10] YANG Wei-jia, QIAO San-zheng. A novel anonymization algorithm: privacy protection and knowledge preservation[J]. Expert Systems with Applications, 2010, 37(1): 756-766.

[11] XIAO Xiao-kui, TAO Yu-fei. Anatomy: simple and effective privacy preservation[C]//Proc of the 32nd International Conference on Very Large Data Bases. 2006: 139-150.