(p,a)-sensitive k-匿名隐私保护模型

王 茜, 曾子平
(重庆大学 计算机学院, 重庆 400044)

摘 要: 提出了一种(p,a)-sensitive k-匿名模型,将敏感属性根据敏感度进行分组,然后给各分组设置不同的约束,并给出了(p,a)-sensitive K-匿名算法。实验结果表明该方法可以明显地减少隐私泄露,增强了数据发布的安全性。

关键词: 数据发布; 敏感度; K-匿名; 隐私泄露; 分组
中图分类号: TP309 **文献标志码**: A **文章编号**: 1001-3695(2009)06-2177-03
doi:10.3969/j.issn.1001-3695.2009.06.054

(p, a)-sensitive k-anonymity: privacy protection model

WANG Qian, ZENG Zi-ping
(*College of Computer Science, Chongqing University, Chongqing* 400044, *China*)

Abstract: This paper proposed a novel model (p,a)-sensitive k-anonymity. It divided sensitive attributes into groups according to the sensitivity, and set each group with different restriction. Described the corresponding algorithm to implement the idea. The result of the experiments suggests that the new model is able to reduce privacy disclosure apparently and enforce security of data publishing.

Key words: data publishing; sensitivity; k-anonymity; privacy disclosure; group

随着网络信息技术的高速发展, 人类大量个人信息被政府部门、商业机构等存储、发布,这极大激发了各部门从海量数据中挖掘有用信息的需求。但事物往往具有两面性, 当数据挖掘用于公开分析大量的私人信息(如购物习惯、犯罪记录、病史、信用记录等)时,它在为人们提供强大知识发现功能的同时,也对个人隐私带来威胁。虽然独立的数据发布单位分别会采取措施隐藏发布数据中的个人身份标志或某些隐私数据,但是值得注意的是通过在多个公开的数据源间进行连接操作往往会导致意想不到的隐私信息泄露问题。文献[1]中的研究表明,通过 zip code、sex、data of birth 等属性对选民登记表和隐匿了个体标志的医疗信息表进行连接操作,超过 87% 的美国公民的身份均可以被惟一标志。

Sweeney 等人[1]提出 k-匿名模型保护隐私数据不受链接攻击,能够有效地解决标志泄露问题,但是对于敏感属性泄露问题并没有相应的保护机制,并且没有过多地考虑敏感属性的敏感度问题,无法达到保护隐私的目的。

1 相关工作

敏感数据发布与共享环境中的个体隐私信息的安全性问题一直是数据隐私研究的热点。k-匿名模型保护隐私数据不受链接攻击,它是指数据表中的每条记录都至少和其他的 $k-1$ 条记录具有相同的准标志符属性取值,但是 k-匿名并没有考虑敏感属性的多样化需求。文献[2]提出在某些情况下 k-匿名并不能保证隐私信息的安全。例如,具有相同准标志符属性取值的所有或大部分记录都具有相同的敏感属性值,恶意的用户还是能够轻易地以高概率推断出隐私信息。因此文献[2]提出了 l-多样性的概念,对匿名化的数据记录中出现频率最高的敏感属性值的个数作出了约束,要求它不大于 $1/l$。P-sensitive k-匿名模型[3]是在 k-匿名的基础上,要求每个等价组中至少要有 p 个不同的敏感属性值,它虽然在一定程度上解决了 k-匿名的敏感属性泄露问题,但是并没有控制敏感属性的出现频率,当 k 比 p 大很多时,就可能出现同一等价组中一个敏感属性的出现频率远大于别的敏感属性,很容易遭受概率攻击。(a,k)-匿名模型[4]通过约束敏感属性值在匿名组中出现的频率小于参数 a 来避免某些敏感值频率过高的情况,增加了敏感值的多样性,防止了一致性攻击,但是它对所有敏感属性都采用同样的约束,不能满足实际要求。文献[5]提出了一种利用经典的空间索引技术来实现 k-匿名的技术,具有很好的扩展性且能很好地支持增量的数据发布。当数据表中准标志符属性的个数较多时,通过概括和隐匿的方法对数据进行 k-匿名化会损失大量信息[6],因此文献[6]提出了一种不基于概括和隐匿的新颖的方法——Anatomy。它通过将原始关系的准标志符属性和敏感属性以两个不同的关系发布,利用它们之间的有损连接来保护隐私数据的安全,并且给出了基本的 Anatomy 算法保证发布的数据满足 l-多样性的要求。

以上提出的敏感数据发布方法在匿名化过程中均没有考虑敏感属性的敏感度问题,在敏感属性具有不同敏感度的数据发布过程中,不能达到有效防止隐私泄露的目的。

2 (p,a)-sensitive k-匿名

2.1 基本概念

假设 T 为原始数据表,T' 为发布的数据表,用于发布或共享的数据表中可以将 T 中的属性分为三类:a)个体身份的标

收稿日期: 2008-09-19; **修回日期**: 2008-10-22

作者简介:王茜(1964-),女,重庆人,副教授,主要研究方向为信息安全、电子商务、远程教育课件工具;曾子平(1983-),男,江西樟树人,硕士研究生,主要研究方向为隐私保护、信息安全(laozeng5614@163.com).

志属性(identifier),如姓名、社会保险号等,通常在数据发布时被隐匿;b)与外部数据源进行连接可标志个体身份的属性,如age、country、zip code 等,叫做准标志符(quasi-identifier,QI);c)包含个体隐私信息的敏感属性,如 disease 等,需要被保护。

定义 1 k-匿名。对于数据表 T,经过匿名化处理得到表 T',T'满足 k-匿名当且仅当在任一等价组中每一条记录至少有 $k-1$ 条记录与它相对应。

k-匿名可以有效防止标志泄露。例如,表 1 为原始数据表,{age,country, zip code}为准标志符,health condition 为敏感属性;经过 2-匿名化后如表 2 所示,即使攻击者知道某条记录在表 2 中,他也无法确定是哪条记录与它相对应。但是,k-匿名无法防止属性泄露问题,因为它没有考虑敏感属性的多样性。例如,对于表 2,假设攻击者能够获得 Rick 的 age = 27,country = Canada,zip code = 14207,且在表 2 中就可以很容易地推断出 Rick 得了 HIV。虽然无法确定是第 1 还是第 2 条记录。

表 1 原始表

ID	age	country	zip code	health condition
1	27	USA	14248	HIV
2	28	Canada	14207	HIV
3	26	USA	14246	cancer
4	25	Canada	14249	cancer
5	41	China	13053	hepatitis
6	48	Japan	13074	phthisis
7	45	India	13064	asthma
8	42	India	13062	heart disease
9	33	USA	14242	flu
10	37	Canada	14204	flu
11	36	Canada	14205	flu
12	35	USA	14248	indigestion

表 2 2-匿名

ID	age	country	zip code	health condition
1	<30	America	142 * *	HIV
2	<30	America	142 * *	HIV
3	<30	America	1424 *	cancer
4	<30	America	1424 *	cancer
5	>40	Asia	130 * *	hepatitis
6	>40	Asia	130 * *	phthisis
7	>40	Asia	130 * *	asthma
8	>40	Asia	130 * *	heart disease
9	3 *	America	1424 *	flu
10	3 *	America	142 * *	flu
11	3 *	America	142 * *	flu
12	3 *	America	1424 *	indigestion

为了解决 k-匿名的属性泄露问题,文献[3]提出了 p-sensitive k-匿名模型。

定义 2 p-sensitive k-匿名。匿名化的数据表 T'满足 p-sensitive k-匿名,当且仅当 T'满足 k-匿名,且在 T'中每个等价组内至少有 p 个不同的敏感属性值。

p-sensitive k-匿名可以在一定程度上解决 k-匿名的属性泄露问题,如表 3 满足 2-sensitive 4-匿名,每个等价组中至少有两个不同的敏感属性值,可以防止一致性攻击。但是在某些情况下,它还是会造成隐私泄露。

表 3 2-sensitive 4-匿名

ID	age	country	zip code	health condition
1	<30	America	142 * *	HIV
2	<30	America	142 * *	HIV
3	<30	America	142 * *	cancer
4	<30	America	142 * *	cancer
5	>40	Asia	130 * *	hepatitis
6	>40	Asia	130 * *	phthisis
7	>40	Asia	130 * *	asthma
8	>40	Asia	130 * *	heart disease
9	3 *	America	142 * *	flu
10	3 *	America	142 * *	flu
11	3 *	America	142 * *	flu
12	3 *	America	142 * *	indigestion

2.2 问题定义

很多时候敏感属性值的敏感度强弱都不相同,如表 1 所示,cancer 与 HIV 的敏感度要远远大于 flu、indigestion 的敏感度。往往人们都不介意别人知道自己得了流感、消化不良等疾病,但是对于 cancer 或者 HIV 病人来说,他们却很怕被人知道,因为这会严重影响到他们今后的生活以及别人对他们的看法,所以对于 cancer 与 HIV 等高敏感度的属性而言,需要提供更强的保护力度。以前的隐私保护模型都没有考虑到敏感属性的敏感度的问题,把它们都统一对待,这显然无法达到保护隐私的目的。可以根据敏感属性的敏感度不同将它们进行分组,例如可以将表 1 中敏感属性 health condition 根据敏感度强弱分为四组(表 4),同一敏感属性组可以包含多个特定的敏感属性值。在这种情况下,不单只是考虑特定敏感属性值,而且要考虑整个敏感属性组的保护,比如,一个人的疾病信息被攻击者知道是属于 top secret 中,虽然不知道具体是 cancer 或 HIV,这也是不能接受的。例如,表 3 是在表 2 的基础上实现的 2-sensitive 4-匿名表,属性值{cancer,cancer,HIV,HIV}都属于 top secret 且在同一个等价组中,这显然也会造成隐私泄露,并且对于高敏感度的敏感属性值而言,它们在同一等价组中出现的频率应该更低。针对上述问题,本文介绍了一种新的更加有效的隐私保护模型(p,a)-sensitive k-匿名模型。

表 4 敏感值分组

敏感组 ID	敏感属性值集合	敏感度
1	HIV, cancer	top secret
2	phthisis, hepatitis	secret
3	heart disease,asthma	less secret
4	flu, indigestion	non secret

2.3 (p,a)-sensitive k-匿名模型

定义 S 为敏感属性值的集合,根据 S 中敏感值的敏感度把 S 分成 m 个组$(S_1,S_2,\cdots,S_m)$。其中 $S=\cup_{i=1}^{m}S_i$,$S_i\cap S_j=\varnothing$ $(i\neq j)$,并且满足 $S_i\leqslant S_j(1\leqslant i<j\leqslant m)$,表示 S_i 比 S_j 更敏感。例如,以表 1 中 health condition 作为敏感属性,则 $S=\{$HIV, cancer,phthisis,hepatitis,heart disease,asthma,flu,indigestion$\}$,根据疾病的敏感度分为了四个组(表 4)。其中 S_1 最敏感而 S_4 最不敏感。

为了对高敏感度的属性值提供更强的保护程度,通过控制高敏感属性值在同一等价组中出现的频率来实现,称为 a 非关联约束。

定义 3 a 非关联约束。对于匿名等价组 E,敏感属性值 s,用(E,s)表示在 E 中敏感属性值为 s 的记录数目,freq$(E,s)=(E,s)/|E|$,a 为一个用户指定的参数,$0\leqslant a\leqslant 1$。组 E 对于敏感属性值 s 来说是满足非关联约束的,如果 freq$(E,s)\leqslant a$。其中,a 必须满足两个条件:a)a 必须不小于敏感属性值 s 在整个数据集上的频率,否则无法实行匿名化;b)a 不小于每个等价组中 s 的最小频率。

传统的匿名模型对所有敏感属性值都采用同一约束值 a,并没有考虑敏感属性值敏感度强弱的问题,显然不能满足实际要求,所以需要定义一个新的约束:敏感组非关联约束 a_{S_i}。

定义 4 敏感组非关联约束。定义 S 为敏感属性值的集合,$S_1,S_2,\cdots,S_m$ 为根据敏感度的一个分组,$a_{S_1},a_{S_2},\cdots,a_{S_m}$ 分别为各个敏感组 a 非关联约束。其中 $a_{S_1}<a_{S_2}<\cdots<a_{S_m}$,并且 a_{S_i}不小于 S_i 中任一敏感属性值 s 的出现频率。

考虑两种极端情况:

a)$a_{S_i}=1$,必然满足,表示 S_i 中属性值不具有敏感性。

b)$a_{S_i}=0$,表示 S_i 中的属性值不想与个体有任何相关,即 S_i 中属性值是不能发布的。

定义 5 (p,a)-sensitive k-匿名。对于数据表 T,经过匿名化得到 T',T'满足 k-匿名且每个等价组中有 P 个不同的敏

感属性组,并且等价组内任一敏感属性值满足敏感组非关联约束,则称 T' 满足 (p, a)-sensitive k-匿名。

例如,表 5 满足 (p, a)-sensitive k-匿名,当取 $k=4, a_{S_1}=0.4, a_{S_2}=0.6, a_{S_3}=0.7, a_{S_4}=0.9$。相对于表 3,本文的模型可以克服 p-sensitive k-匿名的缺点,并能极大地减少隐私泄露的可能性。

表 5　$(2,a)$-sensitive k-匿名

ID	age	country	zip code	health condition	敏感组 ID
1	<40	America	1424 *	HIV	1
2	<40	America	1424 *	cancer	1
3	<40	America	1424 *	flu	4
4	<40	America	1424 *	indigestion	4
5	>40	Asia	130 * *	hepatitis	2
6	>40	Asia	130 * *	phthisis	2
7	>40	Asia	130 * *	asthma	3
8	>40	Asia	130 * *	heart disease	3
9	<40	America	142 *	HIV	1
10	<40	America	142 *	cancer	1
11	<40	America	142 *	flu	4
12	<40	America	142 *	flu	4

2.4　算法

这一节介绍实现 (p,a)-sensitive k-匿名的算法。首先用文献[3]使用的最小 p-sensitive k-匿名算法产生一个数据表 T';然后判断 T' 是否满足 (p,a)-sensitive k-匿名,满足则直接输出,如果不满足,则在等价组之间交换记录直到满足每个等价组内 $c \geqslant p$(c 为不同敏感组的个数)且 $\text{freq}(E, s_i) \leqslant a_{S_j}$,其中 $s_i \in S_j$;最后概化产生满足条件的数据表 T^*,输出。算法描述如下:

输入:数据集 T,敏感属性分组 $D=(S_1, S_2, \cdots, S_m)$,用户定义 a_{S_i}, $p, k(2 \leqslant p \leqslant k)$;
输出:满足 (p,a)-sensitive k-匿名的数据集 T^*。
过程:
产生满足 p-sensitive k-匿名数据集 T'
while not($c \geqslant p$ and $\text{freq}(E, s_i) \leqslant a_{S_j}$)
　交换等价组之间的记录
end while
概化形成新的等价组
return T^*

3　实验

实验所使用的数据集为 UCI 机器学习数据库[7]中的 adult 数据集,该数据集在数据匿名化研究中被广泛使用,已成为该领域事实上的标准测试数据集。Adult 数据集包括部分美国人口普查数据。本文取 age、work class、race、gender 作为准标志符,并增加一列 health condition 作为敏感属性,取值为{HIV, cancer, phthisis, hepatitis, heart disease, asthma, flu, indigestion},分组如表 4 所示。随机选取 500 条记录并把 health condition 的属性值随机插入表中。不同敏感属性值的频率和敏感组约束如表 6 所示。

表 6　敏感属性值频率和组约束

敏感属性值	频率	敏感组	a
HIV	0.018	1	0.4
cancer	0.056	1	
phthisis	0.112	2	0.5
hepatitis	0.124	2	
heart disease	0.105	3	0.7
asthma	0.115	3	
flu	0.213	4	0.9
indigestion	0.257	4	

硬件环境:Intel Pentium IV 2.8 GHz CPU,1 GB RAM。

操作系统:Microsoft Windows XP。

编程环境:Eclipse + Hibernate + Microsoft SQL Server 2000 。

3.1　敏感属性泄露分析

用本文提出的模型与 p-sensitive k-匿名模型匿名化数据集根据 k 的不同产生的敏感属性泄露进行对比。实验表明本模型可以显著地减少敏感属性的泄露,如表 7 所示。

表 7　两种模型下的敏感属性泄露

k-匿名	模型	敏感属性泄露个数	k-匿名	模型	敏感属性泄露个数
3-匿名	2-sensitive	34	4-匿名	2-sensitive	40
	$(2,a)$-sensitive	4		$(2,a)$-sensitive	6
	3-sensitive	20		3-sensitive	30
	$(3,a)$-sensitive	3		$(3,a)$-sensitive	2

3.2　执行时间分析

当 k 越小时(因为 $a_{S_1}=0.4$,所以 $k>2$),产生的等价组数就越多,在等价组之间就要进行很多的记录交换,所以消耗的时间就越多;而当 k 较大时,需要交换的记录就会减少,随着 k 的增大,渐渐趋于 p-sensitive 算法的执行时间,如图 1 所示。

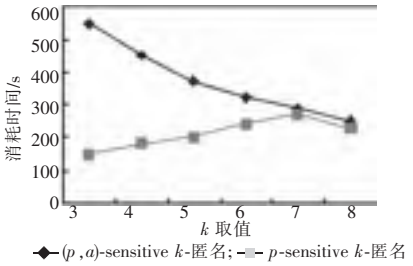

图 1　两种模型下执行时间的对比

4　结束语

隐私信息的安全性是数据发布与共享环境中面临的重要问题。现有的隐私数据发布技术没有考虑敏感属性的敏感度问题,在匿名化过程中将所有敏感属性值都同等对待。针对这一问题,本文提出了一种 (p,a)-sensitive k-匿名隐私保护模型,将敏感属性根据敏感度不同进行分组,并且通过给各敏感组设定不同的约束来提供不同的保护程度。通过实验表明该方法可以有效地防止隐私泄露。

参考文献:

[1] SWEENEY L. K-anonymity: a model for protecting privacy[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 557-570.

[2] MACHANAVAJJHALA A, GEHRKE J, KIFER D, *et al*. L-diversity: privacy beyond k-anonymity[C]//Proc of the 22nd International Conference on Data Engineering. New York: ACM Press, 2006: 24-35.

[3] TRAIAN T M, BINDU V. Privacy protection: p-sensitive k-anonymity property[C]//Proc of the 22nd International Conference on Data Engineering. New York: ACM Press, 2006: 94.

[4] WONG R C, LI Jin-yong, FU A W, *et al*. (a,k)-anonymity: an enhanced k-anonymity model for privacy preserving[C]//Proc of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2006: 754-759.

[5] LWUCHUKWU T, NAUGHTON J. K-anonymization as spatial indexing: toward scalable and incremental anonymization[C]//Proc of the 33rd International Conference on Very Large Data Bases. [S. l.]: VLDB Endowment, 2007: 746-757.

(下转第 2183 页)

定可以计算出相同的会话密钥,而且恶意成员无法欺骗其他成员生成不同的会话密钥。

证明　如果成员都是诚实的,那么每个成员都成功验证前一个成员的中间数据是可靠的,用选取的私钥和前一个成员发送给他的数据进行计算得到他的中间数据并发送给下一个成员,这些中间数值都是可靠的,直到下标最大的成员收到前一个成员的中间数据,并将他计算的中间数据广播给其他成员,其他成员验证数据的可靠性后用中间数据中的有用数值结合自己的私钥可准确地计算出最终会话密钥。

定理 2　一个被动的攻击者得不到会话密钥的任何信息。

证明　被动攻击者试图通过窃听信道中传输的数据来掌握关于会话密钥的信息。本文通过证明攻击者在公开信道窃取的成员 U 的信息可以在不知道成员私钥的情况下仿真来证明攻击者得不到成员 U 的任何信息。

本文用两组随机量来模拟攻击者看到的值和仿真的值,如果不存在概率多项式时间算法能够区分这两组值,这两组值是计算不可分的。由于仿真值是在不知道成员私钥的情况下构造的,并且和真值是计算不可分的,攻击者得不到任何与成员秘密有关的信息。需要一个假设来证明仿真值与真值是计算不可分的,即 G-DDH 困难问题。得证。

定理 3　在随机预言机模型的假设下,协议抗主动攻击。

证明　主动攻击者可以自由选择子会话密钥,并且顺利完成子密钥加密过程;然后,他必须假冒合法用户对消息签名才能通过下一个成员的验证。假设 hash 函数的确是一个随机预言机模型,那么容易证明,如果攻击者可以伪造用户的签名,则该用户的私钥可以被计算出来,而这是不可能的。

定理 4　新加入群的成员不能获悉之前群的会话密钥;离开群的成员不能获悉之后群的会话密钥。

证明　成员加入群时,他所看到的群管理员生成的中间数值已经被群管理者加入随机数更新,所以他不可能得知以前群的会话密钥。同样,在成员离开群时,群管理者将选取新随机数更新他的中间信息广播给其余成员从而更新了会话密钥,因此他同样不能得知以后群的会话密钥。

定理 5　协议满足密钥独立性。

证明　由定理 4,新加入群的成员不能获悉之前群的会话密钥,满足前向保密性;离开群的成员不能获悉之后群的会话密钥,满足后向保密性。因此,满足密钥独立性。

定理 6　即使群成员的长期密钥不慎泄露,利用这些长期密钥生成的会话密钥仍然是安全的。

证明　在协议中,长期密钥用来加密中间数据流,而成员随机选取的私钥在执行下一次协议时会更新。因此,即使长期密钥泄露,这些私钥仍然是安全的,攻击者不可能得到会话密钥。

3　结束语

本文提出的基于 n 叉树的密钥协商方案适合在大群组密钥协商中进一步减少轮数以减少计算复杂度。总之,树在密钥协商中的作用很早就提出,因此,在密钥协商中要多加利用。本文的方案可以看做一个基于树的密钥协商的一个扩展和延伸,根据实际情况可以将 n 量化来加以使用。

参考文献:

[1] DIFFIE W, HELLMAN M. New directions in cryptography[J]. IEEE Trans on Information Theory, 1976, IT-22(6): 644-654.

[2] BRESSON E, CHEVASSUT O, POINTCHEVAL D, *et al*. Provably authenticated group Diffie-Hellman key exchange[C]//Proc of the 8th ACM Conference on Computer and Communications Security. New York: ACM Press, 2001: 255-264.

[3] BRESSON E, CHEVASSUT O, POINTCHEVAL D. Provably authenticated group Diffie-Hellman key exchange: the dynamic case[C]// Proc of the 7th International Conference on the Theory and Application of Cryptology. Berlin: Springer, 2001: 290-309.

[4] BRESSON E, CHEVASSUT O, POINTCHEVAL D. Dynamic group Diffie-Hellman key exchange under standard assumptions[C]//Proc of the 8th International Conference on the Theory and Application of Cryptographic Techniques. London: Springer-Verlag, 2002: 321-336.

[5] BARUA R, DUTTA R, SARKAR P. Extending Joux's protocol to multi party key agreement[C]//Proc of the 4th International Conference on Cyrptology. Berlin: Springer, 2003: 33-60.

[6] DUTTA R, BARUA R, SARKAR P. Provably secure authenticated tree based group key agreement[C]//Proc of the 6th International Conference on Information and Communications Security. Berlin: Springer, 2004: 92-104.

[7] DUTTA R, BARUA R. Dynamic group key agreement in tree-based setting[C]//Proc of the 10th Australasian Conference on Information Security and Privacy. Berlin: Springer, 2005: 101-112.

[8] DUTTA R, BARUA R. Overview of key agreement protocols, report 2005/289[R/OL]. (2005). http://eprint.iacr.org/2005/289.ps.

[9] GÜNTHER C G. An identity-based key-exchange protocol[C]//Proc of Workshop on the Theory and Application of Cryptographic Techniques. New York: Springer-Verlag, 1990: 29-37.

(上接第 2179 页)

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

[7] HETTICH S, BLAKE C L, MERZ C J. UCI repository of machine learning databases[EB/OL]. (1998). http://www.ics.uci.edu/~mlearn/MLRepository.html.

[8] BAYARDO R, AGRAWAL R. Data privacy through optimal k-anonymity[C]//Proc of the 21st International Conference on Data Engineering. Washington DC: IEEE Computer Society, 2005: 217-228.

[9] LI Ning-hui, LI Tian-cheng, SURESH V. T-closeness: privacy beyond k-anonymity and L-diversity[C]//Proc of the 23rd International Conference on Data Engineering. 2007: 106-115.

[10] SWEENEY L. Achieving k-anonymity privacy protection using generalization and suppression[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 571-588.

[11] XU Jian, WANG Wei, PEI Jian, *et al*. Utility-based anonymization using local recoding[C]//Proc of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2006: 785-790.

[12] LEFEVRE K, DEWITT D J, RAMAKRISHNAN R. Incognito: efficient full-domain K-anonymity[C]//Proc of ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 2005: 49-60.