基于 DMSA 算法的多敏感属性数据重发布隐私保护新策略①

左苏楠, 卞艺杰, 吴 慧

(河海大学 商学院, 南京 211100)

摘 要: 针对多敏感属性数据重发布面临的链接攻击、背景知识攻击的威胁, 本文首先提出了敏感属性更新集和同一等价敏感组的概念, 然后对常见的数据重发布情况, 提出了基于 DMSA 算法的数据重发布新策略, 最后对其新策略进行了具体的实例分析, 验证了该策略发布的安全性并得出其隐匿率和附加信息损失度的值都很低, 从而验证出匿名发布后的数据可用性较高, 且具有良好的隐私保护效果.

关键词: 隐私保护; 数据重发布; 多敏感属性; DMSA 算法; 匿名化

Some New Strategies of Data Redistribution of Multisensitive Attributes for Privacy Protection Based on DMSA Algorithm

ZUO Su-Nan, BIAN Yi-Jie, WU Hui

(Business School, Hohai University, Nanjing 211100, China)

Abstract: According to the threats of link attack and background knowledge attack in redistribution of data of multi sensitive attributes, this paper firstly puts forward the concept of update set of sensitive attribute and the same equivalence of the sensitive group. Then, for some common situations of data redistribution, it puts forward some new strategies of data redistribution based on DMSA algorithm. Finally, it is proved by experiments that it has low occult rate and loss degree of additional information. The result also indicates data has high availability and good privacy after released with this new strategy.

Key words: privacy protection; data re-publication; multiple sensitive attributes; DMSA algorithm; anonymous

在大数据时代下, 数据动态更新越来越频繁, 静态数据发布已经不符合人们对数据共享的要求, 因此常常需要数据重发布即数据多次发布. 目前大部分的文献都是针对数据发布的匿名隐私保护做相关研究, 而对于数据重发布隐私保护研究相对较少. 文献[1]基于 l-diversity 策略研究数据添加和删除这两种情况的数据重发布隐私保护问题, 提出了一种数据重发布的隐私保护匿名策略, 即 m-invariance 的泛化策略. 该 m-invariance 规则的不足之处在于无法支持数据属性修改, 且处理后的数据信息损失非常大. 文献[2]基于 k-anonymity 模型研究数据重发布的隐私保护问题. 其不足之处在于只研究了其记录添加的情况, 而且并没有对添加的数据记录根据不同的情况进行分析. 文献[3]基于 l-diversity 规则研究数据重发布的隐私保护问题, 提出了 L-Scarcity 策略, 并且引用群分组(Cohort-Based Partition)和角色分组(Role-Based Partition)的思想来符合 L-Scarcity 策略.文献[4,5,6] 提出了局部保护策略(Localized Guarantee), 该方法规定攻击者对特定时刻的匿名公开数据进行攻击, 得到某一个体的敏感信息的概率不能高于某一个定值。国内数据重新发布的隐私保护研究起步比较晚, 文献[7]针对持续性增长的数据发布做了探讨, 要求在匿名过程中, 必须使匿名数据满足 k-anonymity 和 l-diversity, 但是该文献只研究了数据添加的情况, 并没有对数据更新和数据删除做相关的研究。文献[8]提出了基于 R 树的 k-anonymity 模型, 引入聚类多路分组算法, 设计了

① 收稿时间:2015-05-30;收到修改稿时间:2015-08-17

R 树 k-anonymity 模型。文献[9]针对敏感属性产生邻近攻击的威胁, 引入统一概化的概念, 针对新数据添加的情况, 对敏感数值重新发布设计了一种新模型。文献[10]等学者沿用 m-invariance 策略, 对随着时间固定不变的敏感值做了详细的研究, 设计出一种针对固定不变敏感值的重新发布策略, 并应用 k-means 聚类算法来实现该策略。文献[11]基于 m- invariance 规则为基础并进一步扩展, 设计了 m- inclusion 策略, 引入聚类的理念来是实现该规则, 整个匿名化过程信息损失度非常少, 而且使数据的插入更灵活, 但是它并未对数据的更新操作做探讨。文献[12]基于事务型 k-anonymity 原则提出了一种面向动态集值属性数据重发布的隐私保护模型, 通过保持记录中敏感元素在更新过程中的连续性和多样性阻止其被揭露。

综上所述, 数据发布面临的隐私保护问题得到了国内外学者的高度重视, 对于数据单次发布的研究目前已取得了很大的研究进展, 但是对于数据多次发布的研究国内仍然处于起步阶段, 尤其对于多敏感属性的数据重新发布. 目前, 数据重新发布的研究对象大多是单维敏感属性的数据表, 但是在大数据时代中, 多维敏感属性的数据重新发布问题较为频繁, 如果把目前已有的数据重新发布模型简单套用在多维敏感属性数据重发布中, 会造成大量的隐私信息泄露, 且存在数据重发布后数据可用性较低的问题, 因此对于这些问题亟需进一步的研究.

1 多敏感属性数据重发布概念

在大数据时代下, 数据动态更新越来越频繁, 静态数据发布已经不符合人们对数据共享的要求, 因此常常需要数据重发布即数据多次发布. 而多敏感属性的数据重发布是指所发布的数据表中含有多个敏感属性, 且该数据表在时刻 t_1 已经经过匿名处理发布了一次, 经过一段时间, 原始数据表可能出现新记录的添加, 或者原数据表中的数据记录被修改、或者原数据表中的数据记录被删除, 因此必须在时刻 t_2 处对新的数据表进行匿名重发布. 如果发布者还是按照原来的静态模型进行匿名发布, 攻击者就会根据多个发布版本之间的数据对这几个版本进行有效链接, 经过推理就会使用户的个人隐私被泄露.

2 基于DMSA算法的多敏感属性数据重发布新策略

2.1 敏感属性值更新候选集

同一条数据记录的修改分为以下两种: 一种是敏感属性值的修改, 一种是准标识符属性的修改. 其中敏感属性值的修改又分为两种, 一种是同一记录修改前后的敏感属性值没有任何关联性, 例如张三的健康状态, 从感冒到骨折. 另一种则是同一记录修改前后的敏感属性值存在一定的关联性, 例如李四的健康状态从感冒到肺炎. 为了抵御攻击者将多个匿名发布版本连接起来, 通过背景知识进行攻击, 导致个人隐私的泄露, 本文提出了敏感属性值更新候选集的概念, 即具有一定关联关系的敏感属性值的集合, 这里的关联关系是指敏感属性值 a, b 的关系, 可以由 a 转化到 b, 或者由 b 转化到 a, 称 a、b 互为候选集. 例如敏感属性疾病中感冒和肺炎这两个值就应该属于同一个敏感属性值更新候选集中.

2.2 IDF 技术

逆文档频率 IDF(Inverse Document Frequency)[13]是信息检索中常用的关键词权重评估技术, 是指一个单词出现的文本频数越小, 它区别于不同类别文本的能力就越大, 如在文档集或语料库中, 出现频率高的关键词比出现频率低的关键词传达的信息需求低, 那么出现频率高的关键词应被分配较低的权重. 根据逆文档频率 IDF 可知, 在多敏感属性数据表 T 中, 某一敏感属性中出现次数较高的敏感属性值的敏感程度要明显低于出现次数较低的敏感属性值的敏感程度, 即敏感属性值出现的次数与敏感属性值的敏感程度成反比. 因此本文引进逆文档频率 IDF 技术来确定各个敏感属性的敏感等级, 如果敏感属性值是数值型属性, 则把数值型属性值映射到对应区间并转化成分类型属性来进行处理. 其公式如下:

$$\mathrm{SA}_{ij-}\deg ree=\frac{c(SA_{ij})}{m} \tag{1}$$

式中 $SA_{ij-}\deg ree$ 表示敏感属性 SA_i 中敏感属性值 SA_{ij} 的敏感度, 其中 m 为多敏感属性数据表 T 中的总记录数, $c(SA_{ij})$ 表示敏感属性值 SA_{ij} 在敏感属性 SA_i 中出现的次数.

2.3 DMSA 算法

输入: 待发布的多敏感属性表 $T\{QI_1,QI_2,\cdots,QI_p,SA_1,SA_2,\cdots,SA_q\}$, 多样性约束

$(l_1,l_2,\cdots,l_q)$，$QT_i(1\le i\le p)$ 为准标识符属性，$SA_j(1\le j\le q)$ 为敏感属性，多敏感属性数据表 T 中有 m 条数据，n 个属性，q 个敏感属性.

输出：准标识符属性表 QIT 和敏感属性表 SAT，两张表通过分组号联接起来.

Step1：根据待发表的多敏感属性表 T 中的敏感数据(数据量要多)计算每个敏感属性 $\{SA_1,SA_2,\cdots,SA_q\}$ 的信息熵值，根据每个敏感属性的信息熵值来确定敏感属性的敏感等级. 假设多敏感属性数据表 T 中对敏感属性进行重新排序 $\{SA'_1,SA'_2,\cdots,SA'_q\}$，其中 SA'_1 的敏感属性敏感最高，SA'_q 的敏感属性敏感等级最低；然后根据 IDF 技术确定各个 SA 的敏感值敏感等级.

Step2：计算多敏感属性表 T 中各个敏感属性的敏感属性值的多样性总数，每个敏感属性值的多样性个数为 $(l'_1,l'_2,\cdots,l'_q)$，根据 $(l'_1,l'_2,\cdots,l'_q)$ 的值来定敏感属性多样性约束 $(l_1,l_2,\cdots,l_q)$，其中 $(l_1\le l'_1,l_2\le l'_2,\cdots,l_q\le l'_q)$.

Step3：引入 Anatomy 技术的思想对多敏感属性表进行分组. 分组如下：

把每一维不同的敏感属性值看成是一个向量，按照敏感属性值的敏感级别进行排列，并将表中敏感属性对应的敏感值所在的第 d 条数据(用符号 t_d 表示)一一映射到每一维敏感属性对应的敏感属性值向量中. 在敏感属性等级最高的敏感属性 SA'_i 中的敏感属性值等级最高的向量中任选一条记录 t_d，然后排除其他敏感属性和 t_d 在同一向量中的所有数据. 由于同一分组中每一维敏感属性值要多样化，并且相近的敏感值不能单独分在一组，所以在敏感属性等级最高的敏感属性 SA'_i 中的敏感属性值等级最高的向量中选出与 t_d 不同的数据记录，然后排除这些数据记录在其他敏感属性的同一向量中的其他数据. 接下来对下一维敏感属性进行操作，从剩下的数据记录中挑选一条记录. 然后以此规律循环，找到符合 $(l_2,l_2,\cdots,l_2)-diversity$ 的分组.

Step4：处理未被顺利分组的记录. 对于未被顺利分组的记录，在不损害现有分组满足 $(l_1,l_2,...,l_q)-diversity$ 模型的假设下，可以把它依次分配到已分好的小组中. 其分配的规则是保证同一分组中每一维敏感属性都满足 l-diversity 模型，且保证没有等级相近的 SA 值分在同一组，即确保在同一分组中 SA 值属于互不相同的敏感等级. 最后，在数据发布时，用抑制或泛化技术隐匿未被顺利分组的数据，本文是把不能分组的 SA 值用*来代替. 经过这四步对原始数据进行匿名化，最后将每一分组 G 中的准标识符属性和多敏感属性分别发布成两张表：QIT 表和 SAT 表，两张表通过分组号进行有损链接，从而达到多敏感属性数据发布隐私保护的目的.

2.4 数据添加时的重发布新策略

在实际应用中，数据的添加一般分为两种，一种是增加的数据记录并不在已经发布的匿名表中即是新数据，另一种则是添加的数据记录有一些已经存在于匿名发布表当中即旧数据. 攻击者将两个匿名发布版本连接起来，根据掌握的背景知识，就会造成隐私泄露.

2.4.1 新数据添加时的重发布策略

多敏感属性数据表第一次匿名发布时即满足差异化的多样性模型，若按照 DMSA 算法来进行匿名分组，那么第一次发布的匿名数据是安全可行的. 如果添加的数据记录都是新的数据记录，并没有存在于以前的匿名版本之中，那就说明新数据与之前匿名发布的版本没有任何的关联，这样攻击者就不能通过将两个匿名版本进行连接找到任何攻击的地方. 在面对新数据的添加时，可以把多敏感属性数据重发布问题简化，看成是多敏感属性数据的数据单次发布问题. 因此本文认为对于新数据的添加，可以采用在满足差异化多敏感属性 $(l_1,l_2,\cdots,l_q)-diversity$ 模型的基础之上，采用 DMSA 算法进行匿名化分组处理，即能保证数据发布的安全性也能减少数据损失.

2.4.2 旧数据添加时的重发布策略

对于添加的数据记录已经存在于第一次匿名发布的数据当中，如果该条记录在上一个版本中属于隐匿处理过的数据，那么就可以按照新记录的添加策略即差异化多敏感属性模型对其进行重发布. 对于添加的数据记录已经存在于第一次匿名发布的数据当中，但是该条记录在上一个版本中不是隐匿处理过的数据，而是存在于分组中，那么在此则不能直接沿用差异化多敏感属性模型对其进行处理，应该找到该条记录所在上一个版本所在的分组，并记录该分组中出现的敏感属性值记录，在新添加的数据记录集中寻找新纪录组成一个分组，该分组必须满足 $(l_1,l_2,\cdots,l_q)-diversity$ 模型且与上一个版本中具有相同的等价敏感属性值分组. 等价敏感属性值分组是指

在不同版本中, 具有相同的敏感属性值分组. 例如表 1 中的敏感属性分组号 1 和表 2 中的敏感属性分组号 5 就是具有相同的敏感属性值. 我们称表 1 中的敏感属性分组号 1 和表 2 中的敏感属性分组号 5 称为互为等价敏感组. 我们根据表 1 和表 2 可以得知, 按照此策略发布的数据是安全的, 假设即使攻击者知道王毅(39, M, 211010)在时段和时段去过医院, 可是并不能推出王毅所得了什么病, 因此按照此策略进行数据重发布是安全的. 但是在实际情况中, 有可能并不能找到可以按照上一个版本的等价敏感组进行分组的新记录, 这时可以根据具体情况进行分析, 可以采取对此记录进行匿名处理, 或者推迟发布该条数据记录.

表 1 匿名数据表 $T_{t_i}^* - QIT$

Age	Sex	Zip	Gid
25	M	211000	1
25	F	211003	1
39	M	211010	1

表 2 匿名数据表 $T_{t_j}^*$ -QIT

Age	Sex	Zip	Gid
21	F	211012	5
22	F	211004	5
39	M	211010	5

T_{t_i} 时刻匿名发布的 SAT*

Gid	Disease	Physician
1	Fracture	Jesse
1	Cancer	Gery
1	Asthma	Hali

T_{t_j} 时刻匿名发布的 SAT*

Gid	Disease	Physician
5	Fracture	Jesse
5	Cancer	Gery
5	Asthma	Hali

2.5 数据修改时的重发布新策略

在实际应用中, 数据的修改根据表中属性可以划分为两种, 一种是该条记录的准标识符属性发生变化, 例如该条记录的家庭住址发生了变化, 又或者该条记录的邮政编码发生了变化; 另一种是该条记录的敏感属性值的修改, 例如疾病从感冒恶化到肺炎.

2.5.1 准标识符属性修改时的重发布策略

对于准标识符属性数据发生改变时, 最关键的就是要把这条数据记录在两个匿名版本之间的链接切断, 当这条修改数据在两个版本之间具有相同的等价敏感属性分组, 那么即使攻击者在拥有背景知识的基础上, 也不能推理出该条记录的个人隐私信息, 因为每一个匿名版本都满足 $(l_1, l_2, \cdots, l_q) - diversity$ 模型, 攻击者攻击的概率为 $\frac{1}{\mathrm{I}q}$, 而且应用 DMSA 算法进行分组, 附加信息损失度都非常低, 且匿名率非常的低, 是非常可行的匿名算法, 当攻击者通过连接找到两组分组, 并确定某个个体就在这两个分组号之中, 由于两个分组号具有相同的敏感属性值, 且同一分组中每一维敏感属性值都不相同, 且敏感属性等级不相近, 攻击者很难找到某个个体的具体信息. 所以, 在两个版本之间, 重新找到准标识符属性修改之后的数据记录的分组与原来的数据记录具有相同的等价敏感属性分组是关键, 类似的, 可以把他看成是一个旧记录的添加来处理, 处理方式如 2.3.2.

2.5.2 敏感属性修改时的重发布策略

敏感属性修改之后, 如果继续按照上一匿名发布版本中的同一等价敏感属性分组策略进行匿名处理的话, 攻击者在拥有背景知识的基础上, 根据敏感属性值具有的关联进行推理, 就可以推出该用户在两个发布版本中各自拥有的详细信息记录. 所以本文引入敏感属性更新候选集的概念, 选择同一等价敏感属性分组的敏感属性更新候选集中的记录, 从而抵制攻击者根据敏感属性值之间存在的关联推出个人隐私数据. 如果选择同一等价敏感属性分组的敏感属性更新候选集中的记录, 那么推理出来的关系记录就不唯一, 那么攻击者在没有更多的背景知识的前提下, 是很难确定该记录所在各个分组的每一条敏感属性值.

当敏感属性记录发生改变时, 首先找到该条记录出现的匿名发布版本的分组号, 并记录该分组号的敏感属性; 然后确定同一等价敏感属性分组中的敏感属性更新候选集; 最后在添加的新数据记录集中寻找该条记录在上一个版本中出现的同一等价敏感属性分组中的敏感属性更新候选集中选择新的数据记录, 按照 DMSA 算法来构建同一分组, 从而满足差异化 $(l_1, l_2, \cdots, l_q) - diversity$ 模型.

2.6 数据删除时的重发布新策略

当出现数据删除时, m-不变性策略是采用添加伪记录的策略, 这种方法会造成很多的信息干扰, 影响数据的有效性. 本文针对 m-不变性对于数据删除的重发布不足, 提出当匿名发布之后的数据删除时, 可以

在需要添加的新记录中或者在隐匿的数据中挑选具有相同敏感属性的记录放到该分组当中, 或者把具有相同准标识符属性的数据放到该分组当中, 如果实在是没有找到这样的记录, 则推迟发布.

3　基于DMSA算法的多敏感属性重发布新策略分析

接下来使用具体的实例, 对提出的数据重发布新策略进行分析. 表 3 为多敏感属性原始数据表 T_{t_n}, 表 4 为多敏感属性数据发布表 T_{t_n} 按照DMSA算法进行匿名发布之后的表 $T_{t_n}^*$. 为了读者更好的理解, 文中表格的内容形式采用中文显示, 表 5 为多敏感属性数据表发生变化的数据, 相对于表 3, 表 5 中新添加的数据记录有 8 条, 分别为 $\{t_1,t_3,t_5,t_6,t_7,t_{10},t_{11},t_{13}\}$, 其中旧数据的添加记录有 1 条 $\{t_2\}$, 数据修改的记录有 2 条, 分别为 $\{t_4,t_8\}$, 删除的数据记录有两条 $\{t_9,t_{12}\}$, 文化程度的总类分为博士, 硕士, 本科, 大专, 中专, 高中, 初中, 小学, 文盲与半文盲, 学历的敏感属性更新集如表 6 所示. 按照 DMSA 算法的多敏感属性数据重发布新策略, 对数据添加、数据修改、数据删除时进行匿名发布之后的表如表 7 所示, 当表 7 发布时, 将表 3 中对应的数据记录 t_{12} 删除, 攻击者通过背景知识来链接表 6 和表 7, 攻击的概率为 $\frac{1}{3}*\frac{1}{3}=\frac{1}{9}$, 可以满足发布的安全性, 且根据附加信息损失度的公式 $AIL=\frac{1}{q}\Sigma_{1\le i\le q}\Sigma_{1\le j\le h}\frac{\left|SA_iG_j\right|-l_i}{hl_i}$ 计算的附加信息损失度为 $\frac{1}{2}*\frac{4-3}{3*3}=\frac{1}{18}$, 根据隐匿率的公式 $SuppRatio=\frac{\Sigma_{i=1}^{q}n_{SA_i}}{q|T|}$ 计算匿名率为 0. 综上所述, 本文提出的DMSA算法的多敏感属性数据重发布新策略具有较高的发布安全性, 且信息损失度较小.

表 3　多敏感属性原始数据表

序号	姓名	年龄	性别	邮政编码	学历	疾病
t_1	张忆温	38	女	211001	小学	艾滋
t_2	李化隆	27	男	211013	大专	癌症
t_3	沈媛媛	50	女	211021	高中	感冒
t_4	孙敏郝	63	男	231002	中专	肺炎
t_5	马小丽	27	男	211012	大学	胃炎
t_6	徐军辉	58	女	211022	大专	肺炎
t_7	王雯纹	30	女	211142	中专	感冒
t_8	张尔霞	33	男	212332	大学	胃炎
t_9	冯小飞	34	女	211344	中专	感冒
t_{10}	罗青青	37	女	211552	大学	胃炎
t_{11}	胡明明	41	男	222112	大专	癌症

表 4　多敏感属性匿名发布表 $T_{t_n}^*$ -QIT

年龄	性别	邮政编码	分组号
38	女	211001	1
27	男	211013	2
50	女	211021	1
63	男	231002	3
27	男	211012	1
58	女	211022	1
30	女	211142	2
33	男	212332	2
34	女	211344	
37	女	211552	3
41	男	222112	3
22	男	213112	

$T_{t_n}^*$ -SAT

学历	疾病	
1 小学	艾滋	1
3 高中	感冒	1
5 大学	胃炎	1
6 大专	肺炎	1
2 大专	癌症	2
7 大学	胃炎	2
8 中专	感冒	2
10 大学	胃炎	3
11 大专	癌症	3
4 中专	肺炎	3
*	感冒	
中专	*	

表 5　多敏感属性数据表发生变化的数据

序号	姓名	年龄	性别	邮政编码	学历	疾病	数据变化详情
t_1	赵冬娜	20	女	211131	小学	风湿	新数据的添加
t_2	李化隆	27	男	211013	大专	癌症	旧数据的添加
t_3	李楠木	26	男	233300	大学	胃炎	新数据的添加
t_4	王雯纹	30	女	211142	中专	感冒/肺炎	疾病恶化更新纪录
t_5	钱邻莉	32	女	324909	中专	感冒	新数据的添加
t_6	孙思达	58	男	211022	大专	胃炎	新数据的添加
t_7	李洪波	30	男	211142	大学	胃癌	新数据的添加

t_8	周浩然	33	男	212322	8 中专	感冒	邮政编码发生改变
t_9	孙敏郝	63	男	231002	中专	肺炎	删掉数据
t_{10}	冯志刚	37	男	211552	中专	肺炎	新数据的添加
t_{11}	陈志文	55	男	222112	大专	癌症	新数据记录
t_{12}	汪小虾	22	男	213112	大专	感冒	删除的数据
t_{13}	李海伟	23	男	211100	硕士	骨折	新数据添加

表 6 学历敏感属性值更新集

学历敏感属性值	敏感属性更新集
博士	博士、硕士
硕士	硕士， 博士，本科
本科	本科，硕士，高中
大专	大专、本科
中专	中专、大专
高中	高中、中专
初中	初中、高中
小学	小学、初中
文盲与半文盲	文盲与半文盲

表 7 多敏感属性匿名发布表 $T^*_{t_{n+1}}$ -QIT

年龄	性别	邮政编码	分组号
27	男	211013	1
26	男	233300	1
27	男	211013	1
30	女	211142	2
58	男	211022	2
30	男	211142	2
20	女	211131	3
33	男	212322	3
55	男	222112	3
23	男	211100	

$T^*_{t_{n+1}}$ -SAT

学历	疾病	
大专	癌症	1
大学	胃炎	1
中专	感冒	1
中专	感冒/肺炎	2
大专	胃炎	2
大学	胃癌	2
小学	风湿	3
中专	感冒	3
大专	癌症	3
硕士	骨折	1

4 结语

本文引入了敏感属性更新集和同一等价敏感组的概念; 然后对三种情况, 即数据添加时的重发布问题、数据修改时的重发布问题和数据删除时的重发布问题提出了基于 DMSA 算法的重发布新策略, 最后分别针对多敏感属性表的这三种重发布情况, 进行了具体的实例分析, 证明了该策略发布的安全性以及证明了该策略具有较低的附加信息损失度和匿名率.

参考文献

1 Xiao X, Tao YF. M-invariance: Towards privacy preserving republication of dynamic datasets. Proc. of ACM SIGMOD International Conference on Management of Data. New York. ACM Press. 2007. 689–700.

2 Fung B, Wang K, Fu AWC, et al. Anonymity for continuous data publishing. Proc. of the 11th International Conference on Extending Database Technology. New York. ACM Press. 2008. 264–275.

3 Bu YY, Fu AWC, Wong RCW, et al. Privacy preserving serial data publishing by role composition. Proc. of the VLDB Endowment, 2008, 1(1): 845–856.

4 Iwuchukwu T, DeWitt DJ, Doan A, et al. K-anonymization as spatial indexing: Toward scalable and incremental anonymization. Proc. Int. Conference on Data Engineering. 2007. 1414–1416.

5 Li F, Zhou S. Challenging More Updates: Towards Anonymous Re-publication of Fully Dynamic Datasets. http://arxiv.org/abs/0806.4703. [2008-07-24].

6 Zhang XL, Hong JB. Secure and effective anonymization against re-publication of dynamic datasets. The 2nd International Conference on Computer Engineering and Technology(ICCET). 2010. 399–411.

7 李素伟.增量数据集下的隐私保护技术研究[硕士学位论文].包头:内蒙古科技大学,2010.

8 张晓琳,李猛,李素伟,汤彪,褚燕华.基于 R 树的 k-匿名技术研究.内蒙古科技大学学报,2010,29(4):355–359.

9 于杰.基于动态数据集面向数值敏感属性的隐私保护技术研究[硕士学位论文].包头:内蒙古科技大学,2011.

10 毕红净.动态数据集隐私保护技术研究[硕士学位论文].包头:内蒙古科技大学,2010.

11 于金英.基于动态数据发布的隐私保护研究[硕士学位论文].北京:北京工业大学,2013.

12 武毅,王丹,蒋宗礼.基于事务型 K-Anonymity 的动态集值属性数据重发布隐私保护方法.计算机研究与发展,2013: 248–256.

13 http://baike.baidu.com/view/6219237.htm.