# 隐私保护的数据发布研究

杨高明 杨 静 张健沛

(哈尔滨工程大学计算机科学与技术学院 哈尔滨 150001)

**摘 要** 随着信息技术的发展,个人隐私泄露成为日益严重的问题,因此迫切需要研究防止数据发布中个人隐私的泄露。为此,许多研究者提出不同的方法用以实现隐私保护的数据发布。为总结前人工作,介绍了隐私保护数据发布技术的研究意义和发展历程,阐述了本领域研究过程中的背景攻击模型和隐私模型,深入分析了用已有的概化/隐匿方法和聚类方法实现匿名数据发布技术,总结了匿名质量有关的信息度量标准,同时探讨了数据更新引起的增量数据发布方法和高维数据、移动数据的发布,最后归纳了目前研究中的问题并展望了本领域进一步的研究趋势。

**关键词** 隐私保护,数据发布,k-匿名,概化,信息度量

**中图法分类号** TP309.2 **文献标识码** A

## Research on Data Publishing of Privacy Preserving

YANG Gao-ming YANG Jing ZHANG Jian-pei

(College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China)

**Abstract** With the development of information technology, privacy leakage becomes a serious problem, therefore, it is in urgent need to prevent personal privacy disclosure in data publishing. For this reason, many researchers have proposed different ways to achieve data publishing of privacy protection. To sum up the previous work, we introduced research significance of privacy protection data release technology and its development process, described background attack model and privacy model during the study in this field, deeply analysed existing generalization / suppression method and clustering method to achieve anonymity data release, summarized information metrics of related anonymous dada quality, also discussed incremental data release method caused by data update as well as high-dimensional data and mobile data release, finally, looked further research trends in this field.

**Keywords** Privacy preserving, Data publishing, k-anonymity, Generalization, Information metrics

## 1 前言

随着信息技术的发展,政府、企业收集了大量的数字信息,他们需要交换和发布数据。未经处理的详细数据包含个人敏感信息,直接发布这些数据将侵犯个人隐私。隐私保护的数据发布(PPDP)任务是开发一种方法或工具,使发布的数据依然有实际用处且个人隐私得到保护。PPDP 涉及计算机科学、统计学、经济学和社会科学,本文从计算机科学的角度总结 PPDP 的国内外发展现状。

设隐私保护的数据格式表为 $T$(EID, QID, SA, NSA),此处 EID 是可以清楚识别记录业主的属性集合,如姓名、身份证号码等;QID 是可以潜在识别记录业主的属性集合;SA 包含敏感的个人详细信息,如疾病、工资等;NSA 是除了前面 3 种属性以外的其他属性。这 4 个集合是不相交的。大多数数据发布方法假设数据表的每个记录代表一个记录业主。

假设敏感数据必须保留,供数据分析。匿名隐私保护的目的旨在隐藏记录业主的身份和敏感数据,因此必须删除明确的记录业主标识符。即使明确的标识符信息被删除,且每一个属性都不能唯一标识一个记录业主,但是它们的组合(准标识符)经常可以唯一标识个体或者少量的记录业主[1]。

为阻止联接攻击,数据发布者发布一个匿名表 $T'$(QID', NSA, SA),对原始表 $T$ 中的准标识符 QID 采用匿名操作得到 QID'。匿名操作隐藏了一些详细信息,使得记录关于 QID'是不可区分的。所以,如果通过 QID'联接到一个记录,也同时联接到在 QID'上有相同值的全部其他记录,匿名问题就是生成满足给定匿名需求的匿名表 $T'$,并使数据效用尽可能大。

## 2 攻击模型和隐私模型

大部分 PPDP 文献假设攻击者背景知识有限,攻击者可以把从外部表得到的个人记录与发布数据表的记录、敏感属性或者数据表本身联接(linkage),分别称之为记录联接、属性联接和表联接。这 3 种联接均假设攻击者知道受害者的准标识符,记录和属性联接进一步假设攻击者知道受害者的记录在发布的表中,目的是从发布表中搜寻受害者的敏感信息。表联接主要是搜寻受害者记录是否出现在发布表中。如果一

到稿日期:2010-10-11 返修日期:2010-12-14 本文受国家自然科学基金(61073043, 61073041, 60873037),黑龙江省自然科学基金(F200901)资助。

杨高明(1974—),男,博士生,CCF 会员,主要研究方向为隐私保护、数据挖掘,E-mail:ygm868@163.com;杨 静(1962—),女,教授,博士生导师,主要研究方向为数据库与知识库、隐私保护;张健沛(1956—),男,教授,博士生导师,主要研究方向为数据库与知识库、隐私保护。

个数据表可以有效阻止攻击者实现这些联接，就认为这个数据表实现了隐私保护。若发布表提供的信息比攻击者背景知识少，使攻击者在攻击前后置信度有大的改变，则称之为概率攻击。

### 2.1 记录联接

$k$-匿名：为通过 QID 保护记录联接，Samarati 和 Sweeney[1,2] 提出 $k$-匿名的概念。如果一个记录在表中至少有其他 $k-1$ 个记录有相同的 $qid$ 值，则这个数据表满足 $k$-匿名。在 $k$-匿名表中，每个记录都与其他 $k-1$ 个记录在 QID 上不能区分，所以通过 QID 把一个受害者联接到具体的记录的概率最多是 $1/k$。$k$-匿名表可以有效阻止记录联接。例如表 1(b)是表 1(a)使用图 1 的分类树通过概化 QID={Job, Sex, Age}实现的，它在 QID 上有 2 个不同的组，每个组包含至少 2 个记录，所以是 2-匿名表。

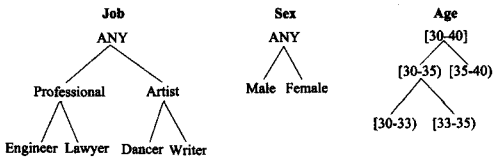


图 1 Job、Sex、Age 的分类树

表 1 各种攻击例子详解表

(a) 患者表

| Job | Sex | Age | Disease |
|---|---|---|---|
| Engineer | Male | 35 | Hepatitis |
| Lawyer | Male | 38 | HIV |
| Writer | Female | 30 | Flu |
| Writer | Female | 30 | HIV |
| Dancer | Female | 30 | HIV |

(b) 2-匿名患者表

| Job | Sex | Age | Disease |
|---|---|---|---|
| Professional | Male | [35-40] | Hepatitis |
| Professional | Male | [35-40] | HIV |
| Artist | Female | [30-35] | Flu |
| Artist | Female | [30-35] | HIV |
| Artist | Female | [30-35] | HIV |

$(X,Y)$-匿名：为研究 $k$-匿名问题，Wang 等[3,4] 提出$(X,Y)$匿名概念，此处 $X,Y$ 是属性的不相交集合。设 $x$ 是 $X$ 上的一个值。$x$ 关于 $Y$ 的匿名表示为 $a_Y(x)$，它是 $Y$ 上 $x$ 发生的不同个数。设 $A_Y(X)=\min\{a_Y(x)\mid x\in X\}$。对指定的整数 $k$，如果 $A_Y(X)\geqslant k$，则 $T$ 满足$(X,Y)$匿名。$(X,Y)$匿名说明 $X$ 上的每个值至少联接到 $Y$ 上 $k$ 个不同值。$(X,Y)$匿名提供了统一、灵活的方法，指定不同类型的隐私需求。如果 $X$ 上的每个值描述一组记录业主(例如 $X$={Job, Sex, Age})，$Y$ 代表敏感属性(如 $Y$={Disease})，则意味着每个组与不同的敏感属性集合相联系，这就很难推导出具体的敏感属性。

多关系 $k$-匿名：大部分匿名工作致力于匿名单个数据表，然而现实生活中的数据库通常包含许多关系表。Nergiz 等[5] 提出了多关系表上的 $k$-匿名隐私保护模型，称为 MultiR $k$-匿名。他们的模型假设关系数据库包含个人详细表 $PT$ 和一系列表 $T_1,\cdots,T_n$。此处 $PT$ 包含个人标识符 Pid 和敏感属性；$T_i(1\leqslant i\leqslant n)$ 包含 QID、敏感属性和外键。广义的隐私概念确保每个记录业主 $o$ 包含在所有表的连接(join)中，且至少 $k-1$ 个其他记录业主与 $o$ 共享相同的 QID。$k$-匿名应用在记录业主层次，而传统 $k$-匿名应用在记录层次。

$k$-匿名、$(X,Y)$匿名和 MultiR $k$ 匿名通过使用相同的 QID 把受害者的记录隐藏在一个大的组中，从而阻止记录联接。然而，如果一个组内的绝大部分记录有相同的敏感属性值，攻击者不需要识别受害者的记录仍然可以关联到他的敏感属性。属性联接可以解决这个问题，下面将进行讨论。

### 2.2 属性联接

$l$-Diversity：Machanavajjhala 等[6] 提出多样性原则来阻止属性联接，称为 $l$-多样性。$l$-多样性要求每个 $qid$ 组至少包含 $l$ 个敏感属性。$l$-多样性不能阻止概率推理攻击，因为在一个组内某些敏感属性很自然地比其他敏感属性频繁。于是出现了两个更强的 $l$-多样性概念，分别是熵多样性和递归$(c,l)$-多样性。如果数据表满足下面条件，则是熵多样性的。

$$-\sum_{s\in S}P(qid,s)\log(P(qid,s))\geqslant\log(l) \tag{1}$$

此处 $S$ 是敏感属性，$P(qid,s)$是敏感值 $s$ 的 $qid$ 组记录函数。等式左边称为敏感属性熵，$qid$ 组内均匀的敏感属性值将产生更大的熵值。因此，大的阈值 $l$ 意味着在一个组内推断出某一敏感属性值的几率减少，而不等式并不依靠 log 基的选择。

递归的$(c,l)$-多样性确保频繁值出现得不太频繁，不频繁值出现得不太稀少。设 $m$ 是 $qid$ 组内的敏感值个数，$f_i$ 表示 $qid$ 组内第 $i$ 个最频繁敏感值频率。如果最频繁的敏感值的频率少于 $m-l+1$ 个最少的频繁敏感值的和与数据发布者指定的约束 $c$ 的乘积，即 $f_1<c\sum_{i=l}^{m}f_i$，则 $qid$ 组满足$(c,l)$-多样性。因此，即使攻击者运用背景知识排除了一些受害者的敏感值，剩下的敏感值仍然很难被推导出来。如果所有的组满足$(c,l)$-多样性，则这个数据表满足递归$(c,l)$-多样性。

$(\alpha,k)$-匿名：Wong 等[7] 提出一个与$(X,Y)$-隐私相似的集成隐私模型，称为$(\alpha,k)$-匿名。它要求表 $T$ 的每个 $qid$ 至少共享 $k$ 个记录，且对每个敏感值 $\mathrm{conf}(qid\rightarrow s)\leqslant\alpha$，此处 $k$ 和 $\alpha$ 都是数据发布者指定的阈值，$\mathrm{conf}(qid\rightarrow s)$表示 $qid$ 组内包含 $s$ 的记录百分比。然而，如果敏感属性值倾斜，$(\alpha,k)$-匿名导致大的失真。

$(k,e)$-匿名：大部分 $k$-匿名及其扩展假设敏感属性为分类属性。Zhang 等[8] 提出$(k,e)$-匿名概念来研究敏感属性为数值属性的情况。基本思想是把记录划分为组，每个组在 $e$ 范围内包含至少 $k$ 个不同的敏感值。然而$(k,e)$-匿名忽略范围 $\lambda$ 上的敏感值的分布。如果某些敏感值在子范围 $\lambda$ 内频繁地发生，那么攻击者可以在组内确切地推导子范围，这类属性联接攻击称为近似攻击[9]。Li 等[9] 提出另外一个隐私模型，称为$(\varepsilon,m)$-匿名。给定 $T$ 上的一些数值的敏感值 $s$，这个隐私模型推导$[s-\varepsilon,s+\varepsilon]$的概率至多为 $1/m$。为达到$(k,e)$-匿名，Zhang[8] 提出一个最优化扰动方法来把数据记录分配到组，使组的误差和 $E$ 最小，其中 $E$ 可以用每个组的敏感值范围度量。最优化算法的时间复杂度和空间复杂度都是 $O(n^2)$，其中 $n$ 是数据的记录总数。

$t$-逼近：Li 等[10] 观察到当敏感属性的总分布倾斜时，$l$-多样性不能阻止属性联接攻击。为阻止倾斜攻击，Li 等[10] 提出 $t$-逼近隐私模型，它要求每个组在 QID 上的敏感属性分布逼近整个表的属性分布。$t$-逼近的限制和弱点有：首先对不同的敏感值，缺乏指定不同保护水平的灵活性；其次它使用的衡量函数不适合阻止数值敏感属性的属性联接[9]；最后，增强的 $t$-逼近将极大地降低数据的效用，因为它要求在所有 qid 组上有相同的敏感值分布，这将明显损害 QID 和敏感属性的关联。一个降低损害的方法是放松要求，在增加倾斜攻击危险

的情况下通过调整阈值实现。

### 2.3 表联接

记录联接和属性联接假设受害者的记录在发布的表 $T$ 中。然而某些情况下受害者的记录是否出现在表 $T$ 中已经泄露了受害者的敏感信息。如果攻击者能确切推导出受害者的记录是否出现在发布的数据表中，表联接就泄露了隐私。

$\delta$-出现：为阻止表联接，Nergiz 等[11]提出限制推导任何受害者记录出现的概率边界，具体范围是 $\delta=(\delta_{\min},\delta_{\max})$。给定一个外部发布表 $E$ 和一个隐私表 $T$，$T\subseteq E$，如果对于全部 $t\in E$，$\delta_{\min}\leqslant P(t\in T|T')\leqslant\delta_{\max}$，则概化表 $T'$ 满足 $(\delta_{\min},\delta_{\max})$ 出现。$\delta$-出现可以间接阻止记录或者属性联接，因为如果攻击者至多有 $\delta\%$ 的置信度目标受害者的记录会出现在发布表中，那么成功地联接到受害者记录敏感属性的概率最多为 $\delta\%$。

一个概化表 $T'$ 关于外部表 $E$ 满足 $(\delta_{\min},\delta_{\max})$ 出现，如果对 $t\in E$，$\delta_{\min}\leqslant P(t\in T|T')\leqslant\delta_{\max}$。为达到 $\delta$-出现，Nergiz 等[11]提出两个匿名算法：SPALM 和 MPALM。SPALM 是使用全域单维概化模式的优化算法。Nergiz 等[11]证明了关于全域概化 $\delta$-出现的反单调特性。如果表 $T$ 是 $\delta$-出现的，$T'$ 的概化版本也是 $\delta$-出现的。SPALM 是一个自上而下的细化方法，使用 $\delta$-出现的反单调特性，取有效的剪枝搜索空间。MPALM 是使用多维概化模式的最小算法，其复杂性为 $O(|C||E|\log_2|E|)$，此处 $|C|$ 是隐私表 $T$ 的属性数，$|E|$ 是外部表 $E$ 的记录数。

### 2.4 概率攻击

概率攻击模型的目的是改变攻击者访问了发布数据以后对受害者敏感信息的概率信度。通常这组隐私模型的目标是确保发布前后置信度的差异很小。

$\varepsilon$-差异隐私：Dwork[12]提出了 $\varepsilon$-差异隐私（differential privacy），它确保添加或者删除单个数据库记录不明显影响任何一个分析结果。如果一个记录业主不能把他的真实信息提供给数据发布者，匿名算法结果中不会有明显差异。Dwark[12]证明了 $\varepsilon$-差异隐私在攻击者拥有任意背景知识时都能提供隐私保护，这种保护通过比较记录业主的数据是否在发布的数据库中达到。设 $n$ 是数据的记录数，如果查询数是线性的，达到不同隐私的噪音边界是 $O(\sqrt{n})$，且不同隐私概念能应用到交互查询和非交互查询。

$(d,\gamma)$-隐私：设 $P(r)$ 是受害者记录在检查数据表 $T$ 之前出现的先验概率，$P(r|T)$ 是检查发布表 $T$ 之后受害者记录在数据表 $T$ 出现的后验概率。$(d,\gamma)$-隐私设定先验概率和后验概率差异的边界，提供一个可证明的隐私和信息效用保证。Rastogi 等[13]说明隐私和效用之间的合理平衡仅仅在先验可信度小的情况下才能达到。然而 $(d,\gamma)$-隐私被设计保护 $d$-独立攻击，如果先验可信度 $P(r)$ 满足对所有记录 $P(r)=1$ 或者 $P(r)\leqslant d$，则攻击是 $d$-独立的。$P(r)=1$ 意味着攻击者已经知道 $r$ 在 $T$ 中，不需要在 $r$ 上提供保护。Machanavajjhala 等[14]指出这个 $d$-独立假设实际上可能无效。不同的隐私在比较中不必假设记录是独立的，或者假设攻击者根据概率分布有先验可信度边界。

## 3 信息损失度量

匿名要使保护隐私和数据效用之间达到平衡。一种信息衡量标准是衡量数据的有用性，即衡量匿名表的数据质量。而搜索衡量标准引导匿名算法在每一步寻找包含最大信息（失真最小）的匿名表，通常通过排列一系列的匿名操作，在每一步贪婪执行最好的搜索来达到。

### 3.1 一般目的度量

许多情况下，数据发布者不知道接收者如何分析发布的数据，这与隐私保护的数据挖掘（PPDM）不同，PPDM 假设数据挖掘任务是知道的。例如 PPDP 数据可以发布到 Web 上，而数据接收者可以按照自己的目的分析数据。一个数据衡量标准对一个数据接收者来说是好的而对另一个就不一定也是好的。在这种场景下，一个合理的信息衡量标准度量原始数据和匿名数据之间的相似性，这是最小失真原则[1,2]。在最小失真衡量标准中，惩罚是由于概化或者隐匿一个值的各个实例引起的。例如，概化工程师到专家的 10 个实例会导致 10 个单位的失真，进一步概化这些实例到 ANY_Job 又会导致 10 个单位的失真。这种衡量标准是对单个属性的度量，以前在文献[1,3,15]中用作数据衡量标准和搜索标准。

*InforLoss* 是文献[16]提出的数据信息损失衡量标准，用来捕获概化详细值到一般值的信息损失。$InforLoss(v_g)=(|v_g|-1)/|D_A|$，$|v_g|$ 是 $v_g$ 子孙的域值数，$|D_A|$ 是 $v_g$ 的属性 $A$ 的域值数。这个数据衡量标准要求知道所有原始数据值在分类系统的叶子节点。如果 $v_g$ 是表中原始数据值，$InforLoss(v_g)=0$，总之 $InforLoss(v_g)$ 度量被 $v_g$ 概化的域值部分。例如，概化图 1 中的 Dancer 到 Artist，信息损失为 $InforLoss(Artist)=(2-1)/4=0.25$，对记录 $r$ 的概化损失为

$$InforLoss(r)=\sum_{v_g\in r}(w_i\times InforLoss(v_g)) \tag{2}$$

式中，$w_i$ 是 $v_g$ 的属性 $A$ 指定的属性权重的处罚正常数。概化表 $T$ 的总信息损失为

$$InforLoss(T)=\sum_{t\in T}InforLoss(r) \tag{3}$$

### 3.2 特殊目的衡量标准

如果数据在发布时使用目的已经明确，就要考虑匿名时更好地保留信息。例如，如果数据是为表中目标属性的分类建模而发布，那么在目标属性中区别类标签特征的值不能概化。为研究分类目标，应该用将来实例的分类误差衡量数据失真。由于很多情况下不能得到将来的数据，大部分方法[17,18]用训练数据测量分类精度。研究结果表明[18]，属性的不同组合可以支配有用的分类知识。

Iyengar[17]首先提出分类衡量标准 CM 用以衡量分类训练数据的错误率。他的思想是如果某个记录不在它所属的类中，则隐匿或者概化这个记录到一个组时施加一个惩罚系数。CM 是数据衡量方法，它对训练数据的修改进行惩罚。这没有完全达到分类目标，但实际上比起概化无用的噪音到有用的分类信息要好得多。对分类来说，一个更相关的方法是按照某些启发式方法搜索好的匿名，即这个方法在每一步搜索时排序匿名操作，而不是最优化数据衡量标准。

### 3.3 平衡衡量标准

特殊用途的信息衡量标准的目的是确保给定的数据挖掘任务数据的有用性。得到最大信息的匿名操作也许会失去隐私，以至于没有其他的匿名操作可以执行。平衡衡量标准的思想是在每个匿名操作时考虑隐私和信息要求，也是在两个要求之间确定寻找的平衡。

Fung 等[4,18,19]基于信息和隐私平衡原则提出搜索衡量标准。假设匿名表重复细化一般值到其孩子值，每个细化操作分裂包含一般值的组为许多组，使每个组包含一个孩子值。

每步操作 $s$ 增益信息，减少隐私。增益的信息表示为 $IG(s)$，损失的信息表示为 $PL(s)$。这个搜索衡量标准倾向于细化 $s$，每次损失隐私时最大化信息增益为

$$IGPL(s)=\frac{IG(s)}{PL(s)+1} \tag{4}$$

$IG(s)$和 $PL(s)$的选择依靠信息衡量标准和隐私模型。对于 $k$-匿名，Fung 等[18,19]用全部 $QID_j$ 上包含属性 $s$ 匿名的平均下降度量隐私损失，即

$$PL(s)=\text{avg}\{A(QID_j)-A_s(QID_j)\} \tag{5}$$

式中，$A(QID_j)$表示细化之前 $QID_j$ 的匿名，$A_s(QID_j)$表示细化之后 $QID_j$ 的匿名。一个变种是设置 $PL(s)$ 为 0，最大化但信息增益。细化使信息增益最大化但也损失了隐私。信息和隐私平衡的原则也可以用来选择概化 $g$，以下情况下它是最小的。

$$ILPG(g)=\frac{IL(g)}{PG(g)+1} \tag{6}$$

式中，$IL(g)$表示信息损失，$PG(g)$表示执行 $g$ 时的信息增益。

## 4 匿名实现技术

一般情况下，原始数据表不能满足指定的隐私保护需求，必须在数据表发布之前应用一系列匿名操作，如概化/隐匿、聚类、分解、排序、扰动等。概化/隐匿用 QID 属性的概化值取代具体的值。分解和排序通过分组和打乱 $qid$ 组内的敏感值来分离 QID 和敏感属性的相关性联系。扰动通过添加噪音、聚集值、交换值，产生基于原始表统计特性的人工数据使数据失真。由于分解和排序仅有一篇文献支撑，扰动倾向于数据挖掘，本文重点讨论概化/隐匿和聚类技术。

### 4.1 概化/隐匿

概化隐藏 QID 上的某些详细值。对于分类属性，具体的值被给定的分类树概化值取代。在图 1 中，父节点 Professional 比子节点 Engineer 和 Lawyer 更一般。根节点 ANY_Job 代表 Job 的最一般值。如果给定了一个数值属性区间分类系统，情况与分类属性相似，更普遍的情况是没有为数值属性预定义分类系统。不同的匿名操作对隐私保护、数据效用和搜索空间影响不同，它们都是原始数据一致的表示，只是精度降低了。概化用属性的分类系统的父节点值取代某些具体值，其相反操作为细化。隐匿操作用一个特殊值取代某些值，标志着取代的值没有泄露，隐匿的相反操作为披露。概化总共有 5 种模式。

全域概化模式：在这种模式中，属性中的所有值都概化到分类树的同一个层。例如图 1 中，如果 Lawyer 和 Engineer 都概化到 Professional，那么它也要求概化 Dance 和 Writer 到 Artist。这种模式的搜索空间比其他模式的搜索空间要小，但是要求所有分类树的路径到同一个粒度层，其数据失真最大。

LeFevre 等[20]使用全域概化提出一系列优化的自下而上的概化算法来产生全部可能的 $k$-匿名，这个算法称为 incognito。算法为计算 $qid$ 组探索汇总特性，若 $qid$ 是$\{qid_1,\cdots,qid_c\}$的概化，则$|qid|=\sum_{i=1}^{c}|qid_i|$。汇总特性说明父母组的大小可以由子女组的大小 $|qid_i|$ 的和直接计算得到，意味着所有可能概化组的大小 $|qid_i|$ 可以由自底向上的方式增量计算。这个特性不仅允许计算组的大小有效，而且为进一步概化提供最终条件。

子树概化模式：在这种模式中，非叶子节点的全部孩子都概化或者一个也没有概化。例如图 1 中，如果 Engineer 概化到 Professional，这个模式也要求其他孩子节点如 Lawyer 概化到 Professional，但是 Dancer 和 Writer 作为 Artist 的孩子节点可以仍然不需概化。概化属性是对分类树的剪枝，它是包含根到叶路径上树的值子集[17,19,21]。

$k$-优化[21]是灵活使用子树概化的有效算法，它通过使用集合枚举树为搜索空间建模，可以有效地剪枝非优化匿名表。每个节点代表一个 $k$-匿名的解决方案。算法假设一个全序属性集合，以自上向下的方式检查枚举树，即自最一般的表开始，使用基于差别的衡量标准(Discernibility Metric，DM)和分类衡量标准(CM)检测。当发现枚举树上某个节点的子孙节点全部不能满足全局最优时，则剪枝该节点。与前面讨论的算法不同，$k$-优化使用子树概化和记录隐匿模式。Machanavajjhala 等[6]修改自下而上的匿名[20]，识别最优的 $l$-多样性表。$l$-多样性基于概化特性是非降的概化过程。使用全域和子树概化的 $k$-匿名算法可以扩展到 $l$-多样性算法。

兄弟概化模式：这个模型与子树概化模式相似，唯一不同是某些兄弟可以不被概化。一个父节点值被解释为代表所有缺失的孩子值。例如图 1 中 Engineer 概化到 Professional，Lawyer 仍然保留不被概化。Profession 被解释为除 Lawyer 外其所覆盖的所有孩子。这个模式比子树概化模式产生的失真少，因为它仅需概化违反指定阈值的子节点。

单元概化模式：前面介绍的模式如果一个值被概化，其全部实例也被概化，这种模式称为全局概化。在单元概化中(也称局部概化)，一个值的某些实例可以保持不变，而其他实例被概化。例如在表 1(a)中，Engineer 在第一个记录中被概化为 Professional，而在第二个记录中却保持不变。比起全局概化模式，这种模式更灵活，因此数据失真小[7]。

Wong[7]使用单元概化模式并提出贪婪的自上而下和自下而上算法识别满足$(\alpha,k)$-匿名的最小匿名解决方案。对 incognito 算法进行了扩展，作者既使用了全局记录模式又使用了局部记录模式，并证明了$(\alpha,k)$-匿名是 NP 难度问题。

多维概化：设 $D_i$ 是属性 $A_i$ 的域，像全域概化和子树概化这样的单维概化可由函数 $f_i$ 定义：对 QID 中的每个属性 $A_i$，$D_{Ai}\rightarrow D'$。单个函数 $f$ 定义的多维概化为 $D_{A1}\times\cdots\times D_{An}\rightarrow D'$，这个定义把 $qid=\langle v_1,\cdots,v_n\rangle$概化为 $qid'=\langle u_1,\cdots,u_n\rangle$，对每个 $v_i$，或者 $v_i=u_i$ 或者 $v_i$ 为 $A_i$ 的分类系统上 $u_i$ 的孩子节点。这个模式允许两个 $qid$ 组独立灵活地概化到不同的父节点组，即使这两个 $qid$ 组有相同的值 $v$ 也一样。例如〈Engineer，Male〉可以概化到〈Engineer，ANY_Sex〉，而〈Engineer，Female〉概化到〈Professional，Female〉。概化表包含 Engineer 和 Professional。这个模式比起全域概化和子树概化产生较少的失真，因为它仅需概化违反指定阈值的 $qid$ 组。在这个多维模式中，所有 $qid$ 组中的记录概化到同一个 $qid'$，但是单元概化没有这个限制[22,23]。

LeFever 等[22]提出一个贪婪的自上而下的细化算法用于在多维概化模式下寻找最小 $k$-匿名。这个算法非常类似于自上而下细化算法(TDS)。这两个算法每次都在值 $v$ 上执行一个细化，主要的不同是 TDS 细化所有包含 $v$ 的 $qid$ 组，即细化仅仅在每个 $qid$ 组包含至少 $k$ 个记录才被执行，而多维 $k$-匿名如果每个细化 $qid$ 组包含至少 $k$ 个记录，则执行细化。因此多维概化的匿名数据通常比单维有更好的数据质量。

Xu 等[24]说明使用单元概化能进一步提高数据质量。LeFevre 等[23]提出一系列贪婪算法来满足 $k$-匿名和 $l$-多样性，同时也考虑具体的数据分析任务，如分类建模多目标属性和最小不精确查询回答。他们的自上而下算法使用了多维概化。

隐匿模式有多种。记录隐匿[1,17,20,21]是隐匿整个记录，值隐匿[25]是隐匿表中给定值的每个实例，单元隐匿（局部隐匿）[26]是隐匿表中给定值的某些实例。

总的说来，匿名操作的选择影响匿名表的搜索空间和数据失真。全域概化搜索空间最小但失真最大，局部概化模式搜索空间最大，但失真最小。若采用局部概化模式，相应的数目就会比较大，因为 QID 上每个属性、任何一个子集值都能被概化而剩余的维持不变。

如果一个表满足给定的隐私要求，并在所有满足的表中按照选择信息的度量标准包含大部分信息，则这个表是最优匿名的。找到最优匿名表是 NP 难度的。Meyerson[26]，Aggarwal[27]证明通过单元隐匿、值隐匿、单元概化达到最优 $k$-匿名是 NP 难度问题。Wong[7]证明通过单元概化达到 $(\alpha,k)$ 匿名是 NP 难度的。

### 4.2 聚类方法

概化数据匿名的方法把数据划分成组，聚类的目的是把数据划分成簇。由于组和簇实际上都是数据集合的表示，因此可以使用聚类实现匿名技术。但是隐私保护的匿名数据发布技术要求每个簇最小为 $k$，而数据聚类技术限定聚类数目为 $k$，因此数据挖掘中的聚类技术不能直接应用到 PPDP，必须开发适应 PPDP 的聚类技术。

Aggarwal 等[28]首先提出使用聚类的 PPDP，他们对每个簇发布 3 个特征：每个簇心的准标识符属性值、每个簇的记录数、敏感属性的值集合。文中为最大簇半径定义了 $r$-Gather 问题，目的是最大化簇的最大半径，同时引入一种 $r$ 单元聚类的聚类衡量标准。他们为了排除离群点，允许 $1/\epsilon$ 的记录不参加聚类。为了改进聚类效果，Aggarwal 等又对该方法进行了深入讨论[29]。

王智慧等[30]针对准标识符中的有序属性和无序属性分别给出了数据概化策略。同时，通过考察数据概化前后属性值不确定性程度的变化，量化地定义了数据概化带来的信息损失。在此基础上，将数据匿名问题转化为带特定约束的聚类问题。文中针对 $l$-多样模型，提出了一种基于聚类的数据匿名方法 $l$-clustering，以在实现数据共享时保护敏感属性的匿名，同时降低概化带来的信息损失。

统计学领域研究 PPDP 问题与以上聚类方法类似，称为微聚集方法。这方面的文献综述可以参考文献[31]。

## 5 增量匿名算法

### 5.1 多视图发布

不同的数据接收者可能对数据表的不同属性感兴趣。假设有数据表 $T$(Job,Sex,Age,Race,Disease,Salary)。一个数据接收者对目标属性{Disease}以及属性{Job,Sex,Age}的分类建模感兴趣。另外一个数据接收者对使用{Job,Age,Race}进行聚类分析感兴趣。一种方法是为两个目的发布属性{Job,Sex,Age,Race}的单发行版本。缺点是这两个目的都不需要发布的全部 4 个属性，它也易受到攻击。一个好的方法是为每个数据挖掘目的匿名并发布一个特制的发行。假设两种版本都发布，很有可能数据接收者有权访问它们，这就很难阻止数据接收者互相之间串通。特别是攻击者可能使用这两个视图组成一个清晰的、包含这两个视图中属性的 QID。

Yao 等[32]提出了探测在一系列视图上违反 $k$-匿名的方法，每个视图由投影得到或者由选择查询得到，他们也考虑将其作为先验知识的函数依赖。宋金玲等[33]指出在视图发布过程中，求解准标识符的关键是如何在已发布的视图集合中找出与待发布视图相关的全部视图。并将已发布的视图集合与待发布的视图映射为一个超图，寻找相关视图集问题可被转化为在超图中求解特定节点间的全部通路问题。他们提出了基于超图的相关视图集求解算法，研究了基本表中属性间不存在函数依赖和存在函数依赖两种情况下准标识符的组成结构，归纳出它们的特征，并给出了基于相关视图集的准标识符求解算法。

### 5.2 增量数据发布

当数据不断发生变化时，数据被增量发布。在增量发布模型中[3]，数据发布者有以前的版本 $T_1,\cdots,T_{p-1}$，现在需要发布 $T_p$，其中 $T_i$ 是 $T_{i-1}$ 插入或者删除数据以后的更新版本。这个问题假设同一个个体的全部记录在所有的发布中不变。即使每个发布 $T_1,\cdots,T_p$ 被单独匿名，隐私需求都可能通过比较不同的版本和排除一些可能的受害者敏感值而受到损害。这个问题假设数据动态更新，进一步说，这个问题假设所有的发布共享同一个数据库模式。

Byun 等[34]首先提出一种插入新记录的隐私保护增量匿名发布技术。它保证每个版本满足 $l$-多样性，这要求每个 $qid$ 组包含至少 $l$ 个不同敏感值。Byun 等[34]研究记录插入而不是删除导致的威胁，因此当前发布的 $T_p$ 包含所有以前发布的记录。只有插入记录以后下面两个隐私需求仍然满足，算法才插入一个记录到当前发布的 $T_p$：1) $T_p$ 是 $l$-多样性的；2) 给定任何一个以前发布的 $T_i$ 和当前发布的 $T_p$，至少有 $l$ 个不同的敏感值仍然在记录中代表受害者的记录。这个需求可以通过比较 $T_i$ 和 $T_p$ 中两个 $qid$ 组敏感值的不同来验证。如果两个隐私需求都得到满足，算法倾向于细化 $T_p$ 时尽可能地提高数据质量。如果一些新记录的插入违反任何一个隐私需求，即使概化以后违反了也要延迟到下一个发布才能执行插入操作。这个策略有时会面临运行到没有新数据可以发布的情况，且要求非常大的内存缓存区存储这些延迟的数据记录。

Xiao 等[35]提出一种保护隐私的匿名方法，称为 $m$-不变性(invariance)，在这种增量数据发布模型中研究记录插入和删除的情况。一个序列发布 $T_1,\cdots,T_p$ 如果满足下列两种情况则是 $m$-不变的：1) 任何一个 $T_i$ 上的每个 $qid$ 组至少包含 $m$ 个记录，且所有记录在 $qid$ 上有不同的敏感值；2) 任何一个记录 $r$ 在发布生命期 $[x,y]$ 内 $(1\leqslant x,y\leqslant p)$，$qid_x,\cdots,qid_y$ 有同样的敏感值集合，$qid_x,\cdots,qid_y$ 是 $T_x,\cdots,T_y$ 上包含 $r$ 的概化 $qid$ 组。$m$-不变性的基本原理是：如果记录 $r$ 在 $T_x,\cdots,T_y$ 上被发布，那么所有包含 $r$ 的 $qid$ 组必须拥有同样的敏感值集合。这将确保在所有这样的 $qid$ 组上敏感值的插入不会引起敏感值集合在各个组之间比较。给定一系列 $m$-不变性 $T_1,\cdots,T_{p-1}$。Xiao[35]通过最小添加伪造的数据记录和概化现在发布的 $T_p$ 来维持一系列 $m$-不变性。

其他增量数据发布还包括宋金铃等[36]在详细分析 $k$-匿名数据集更新情况的基础上，根据语义贴近度及元组映射对更新元组在 $k$-匿名数据集中进行定位，再对更新元组进行相

应的操作。她们的算法不仅保证了数据集的 $k$-匿名约束性质，而且保证了 $k$-匿名数据集与原始数据集的一致性。

## 6 匿名其他类型数据

现在的研究显示，发布事务数据、移动对象数据、文本数据也会导致隐私威胁和敏感信息泄露。下面讨论在非关系数据类型上的隐私威胁和它的一些隐私解决方案

### 6.1 高维事务数据

事务数据通常是高维数据，使用传统的 $k$-匿名隐私模型将要求在单个 QID 中包含全部维。由于维灾难[27]，即使 $k$ 值很小，也有许多数据可能概化到最顶端值，以达到 $k$-匿名，这些匿名的数据对于分析来说显然是没有用的。最近有许多研究关注高维数据匿名。Ghinita 等[37]提出了一种排列方法，其总体思想是首先使用最近邻分组事务数据，接着把每个组与不同的敏感值联系。隐私攻击中，攻击者不可能知道目标受害者的全部准标识符属性，因为他得到每个背景知识的代价太大，然而需要合理地限定攻击者在隐私模型中的背景知识边界。

### 6.2 移动对象数据

基于位置的服务(LBS)是提供移动用户的信息服务，它基于用户的具体位置。移动对象数据是时间依赖的、位置依赖的。它产生于高维数据流中，数据量大。Abul 等[38]扩展传统的 $k$-匿名模型来匿名移动对象集合，使同一个时间周期内每个移动对象的路径半径 $\delta$ 内至少有 $k$ 个移动对象。潘晓等[39]也从匿名连续查询角度提出隐私模型和质量模型来均衡隐私保护与服务质量的矛盾，并基于此提出了一种适用于连续查询贪心匿名的算法。

**结束语** 随着信息技术的发展，收集信息变得容易和便捷，数据挖掘技术也在飞速发展，这些都会造成个人隐私泄露，为此需要对个人隐私信息进行保护。本文主要对目前存在的隐私保护技术进行总结，文中主要从攻击模型和隐私模型对目前常见的、使用概化/隐匿方法实现的匿名技术进行总结。就匿名质量问题对信息度量标准进行了总结。匿名实现技术方面对概化进行了分类并从聚类角度对目前实现的匿名技术进行了总结。针对数据的不断变化总结了多视图发布和增量数据发布。最后说明了高维事务数据和移动对象数据的发展现状。

目前许多方面都涉及到隐私保护问题，比如基于位置服务的移动设备、传感器网络、社会网络等的发展对隐私保护提出新的要求。这方面的研究刚刚起步，有待进一步研究、完善其理论和技术。随着数据量的增大，隐私保护海量数据发布或者数据流发布也有进一步研究的必要。准标识符的选择仍然是一个需要继续研究的问题。

## 参 考 文 献


[1] Samarati P. Protecting Respondents' Identities in Microdata Release [J]. IEEE Transactions on Knowledge and Data Engineering, 2001, 13(6): 1010-1027

[2] 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

[3] Wang K, Fung B C M. Anonymizing sequential releases[C]// Proceedings of KDD 2006. Philadelphia, PA, USA: ACM, 2006: 414-423

[4] Fung B C M, Wang K, Chen R, et al. Privacy-preserving data publishing: A survey of recent developments [J]. ACM Comput. Surv., 2010, 42(4): 1-53

[5] Nergiz M E, Clifton C, Nergiz A E. Multirelational $k$-anonymity [C]// Proceedings of ICDE'07. Istanbul, Turkey, 2007: 1417-1421

[6] Machanavajjhala A, Kifer D, Gehrke J, et al. $l$-diversity: Privacy beyond $k$-anonymity[J]. ACM Transactions on Knowledge Discovery from Data, 2007, 1(1)

[7] Wong R, Li J, Fu A, et al. $(\alpha,k)$-anonymity: an enhanced $k$-anonymity model for privacy preserving data publishing[C]// Proceedings of KDD 2006. ACM, 2006: 754-759

[8] Zhang Q, Koudas N, Srivastava D, et al. Aggregate query answering on anonymized tables[C]// Proceedings of ICDE'07. Istanbul, Turkey, 2007: 116-125

[9] Li J, Tao Y, Xiao X. Preservation of proximity privacy in publishing numerical sensitive data [C]// Proc. ACM SIGMOD Int. Conf. Manage. Data. Vancouver, Canada: ACM, 2008: 473-486

[10] Ninghui L, Tiancheng L, Venkatasubramanian S. $t$-Closeness: Privacy beyond $k$-anonymity and $l$-diversity[C]// Proceedings of ICDE'07. 2007: 106-115

[11] Nergiz M E, Atzori M, Clifton C. Hiding the presence of individuals from shared databases[C]// Proc. ACM SIGMOD Int. Conf. Manage. Data. Beijing, China: ACM, 2007: 676

[12] Dwork C, Mcsherry F, Nissim K, et al. Calibrating noise to sensitivity in private data analysis[C]// 3rd Theory of Cryptography Conference (TCC 2006). New York, NY, United States: Springer Verlag, 2006: 265-284

[13] Rastogi V, Suciu D, Hong S. The boundary between privacy and utility in data publishing[C]// Proceedings of the 33rd International Conference on Very Large Data Bases. Vienna, Austria: VLDB Endowment, 2007: 531-542

[14] Machanavajjhala A, Kifer D, Abowd J, et al. Privacy: Theory meets practice on the map[C]// Proceedings of ICDE'08. Cancun, Mexico, 2008: 277-286

[15] Sweeney L. $k$-anonymity: A model for protecting privacy [J]. Int. J. Uncertainty Fuzziness Knowledge Based Syst., 2002, 10 (5): 557-570

[16] Xiao X, Tao Y. Personalized privacy preservation [C]// Proceedings of SIGMOD 2006. Chicago, I L, USA: ACM, 2006: 229-240

[17] Iyengar V S. Transforming data to satisfy privacy constraints [C]// Proc. of KDD 2002. Edmonton, Alta, Canada, 2002: 279-288

[18] Fung B C M, Wang K, Yu P S. Anonymizing Classification Data for Privacy Preservation [J]. IEEE Transactions on Knowledge and Data Engineering, 2007, 19(5): 711-725

[19] Fung B C M, Wang K, Yu P S. Top-Down Specialization for Information and Privacy Preservation [C]// Proceedings of ICDE' 05. IEEE Computer Society, 2005: 205-216

[20] Lefevre K, Dewitt D J, Ramakrishnan R. Incognito: Efficient full-domain $K$-anonymity[C]// Proceedings of SIGMOD 2005. Baltimore, M D, USA: Association for Computing Machinery, 2005: 49-60

[21] Bayardo R J, Agrawal R. Data privacy through optimal $k$-anonymization[C]// Proceedings of ICDE'05. 2005: 217-228

[22] Lefevre K, Dewitt D J, Ramakrishnan R. Mondrian multidimensional $K$-anonymity [C] // Proceedings of ICDE' 06. Atlanta, USA, 2006: 25-35

[23] Lefevre K, Dewitt D J, Ramakrishnan R. Workload-aware anonymization[C]// Proceedings of SIGKDD 2006. Philadelphia, PA,

USA: ACM, 2006: 277-286
[24] Xu J, Wang W, Pei J, et al. Utility-based anonymization using local recoding [C]// Proceedings of SIGKDD 2006. Philadelphia, PA, USA: ACM, 2006: 785-790
[25] Wang K, Fung B C M, Yu P S. Handicapping attacker's confidence: an alternative to k-anonymization [J]. Knowl. Inf. Syst., 2007, 11(3): 345-368
[26] Meyerson A, Williams R. On the complexity of optimal k-anonymity [C]// Proceedings of PODS 2004. ACM, 2004: 223-228
[27] Aggarwal C C. On k-anonymity and the curse of dimensionality [C]// Proceedings of the 31st International Conference on Very Large Data Bases. Trondheim, Norway: VLDB Endowment, 2005: 901-909
[28] Aggarwal G, Feder T, Kenthapadi K, et al. Achieving anonymity via clustering [C] // Proc. ACM SIGACT SIGMOD SIGART Symp Princ Database Syst. Illinois, USA: ACM, 2006: 153-162
[29] Aggarwal G, Panigrahy R, Tom, et al. Achieving anonymity via clustering [J]. ACM Trans. Algorithms, 2010, 6(3): 1-19
[30] 王智慧,许俭,汪卫,等. 一种基于聚类的数据匿名方法[J]. 软件学报,2010,21(4):680-693
[31] 韩建民,岑婷婷,虞慧群. 数据表 k-匿名化的微聚集算法研究[J]. 电子学报,2008,36(10):2023-2029
[32] Yao C, Wang X S, Jajodia S. Checking for k-anonymity violation by views[C]// Proceedings of VLDB 2005. Trondheim, Norway: Association for Computing Machinery, 2005: 910-921
[33] 宋金玲,刘国华,黄立明,等. k-匿名方法中相关视图集和准标识符的求解算法[J]. 计算机研究与发展,2009,46(1):77-88
[34] Byun J, Sohn Y, Bertino E, et al. Secure anonymization for incremental datasets[C]// Proceedings of the 3rd VLDB Workshop on Secure Data Management. Seoul, Korea: Springer Verlag, 2006: 48-63
[35] Xiao X, Tao Y. M-invariance: towards privacy preserving republication of dynamic datasets[C]// Proceedings of SIGMOD 2007. Beijing, China: ACM, 2007: 689-700
[36] 宋金玲,赵威,刘欣,等. k-匿名数据集的增量更新算法[J]. 计算机科学,2010,37(4):146-150
[37] Ghinita G, Tao Y, Kalnis P. On the anonymization of sparse high-dimensional data[C]// Proceedings of ICDE'08. Cancun, Mexico, 2008: 715-724
[38] Abul O, Bonchi F, Nanni M. Never walk alone: Uncertainty for anonymity in moving objects databases[C]// Proceedings of ICDE'0. Cancun, Mexico, 2008: 376-385
[39] 潘晓,郝兴,孟小峰. 基于位置服务中的连续查询隐私保护研究[J]. 计算机研究与发展,2010,47(1):121-129

(上接第 10 页)

[17] Ng H-H, Soh W-S, Motani M. MACA-U: A Media Access Protocol for Underwater Acoustic Networks[C]// Proc of IEEE GLOBECOM'08. New Orleans: IEEE, 2008: 1-5
[18] Chirdchoo N, Soh W-S, Chua K C. MACA-MN: A MACA-based MAC Protocol for Underwater Acoustic Networks with Packet Train for Multiple Neighbors[C]// Proc of IEEE VTC'08. Spring Singapore: IEEE, 2008: 46-50
[19] Syed A A, Ye Wei, Heidemann J. T-Lohi: A New Class of MAC Protocols for Underwater Acoustic Sensor Networks[C]// Proc of IEEE INFOCOM'08. Phoenix: IEEE, 2008: 231-235
[20] Syed A, Ye Wei, Heidemann J. Comparison and Evaluation of the T-Lohi MAC for Underwater Acoustic Sensor Networks [J]. IEEE Journal on Selected Areas in Communications, 2008, 26(9): 1731-1743
[21] Guerra F, Casari P, Zorzi M. World ocean simulation system (WOSS): a simulation tool for underwater networks with realistic propagation modeling[C]// Proc of ACM ENSS'09. California: ACM, 2009: 1-8
[22] Mirza D, Lu Feng, Schurgers C. TB-MAC Efficient MAC-Layer Broadcast for Underwater Sensor Networks [C] // Proc of ISSNIP'09. Melbourne, Australian, 2009: 1-5
[23] Xie Peng, Cui Jun-hong. R-MAC: An Energy-efficient MAC Protocol for Underwater Sensor Networks[C]// Proc of IEEE WASA'07. Chicago: ACM, 2007: 187-198
[24] Ma Yu-tao, Guo Zhong-wen, Feng Yuan, et al. C-MAC: A TDMA-based MAC Protocol for Underwater Acoustic Sensor Networks[C]// Proc of IEEE NSWCTC'09. Wuhan: IEEE, 2009: 728-731
[25] Hsu C-C, Lai K-F, Chou C-F, et al. ST-MAC: Spatial-Temporal MAC Scheduling for Underwater Sensor Networks[C]// Proc of IEEE INFOCOM'09. Rio de Janeiro: IEEE, 2009: 1827 -1835
[26] Syed A, Heidemann J. Time synchronization for high latency acoustic networks[C]// Proc of IEEE INFOCOM'06. Barcelona: IEEE, 2006
[27] Kalofonos D N, Stojanovic M, Proakis J G. Performance of adaptive MC-CDMA detectors in rapidly fading Rayleigh channels [J]. IEEE Transactions on Wireless Communications, 2003, 2(2): 229-239
[28] Stojanovic M. Low complexity OFDM detector for underwater acoustic channels [C]// Proc of IEEE OCEANS'06. Boston: IEEE, 2006: 1-6
[29] Hayajneh M, Khalil I, Gadallah Y. An OFDMA-based MAC protocol for under water acoustic wireless sensor networks[C]// Proc of ACM IWCMC'09. New York: ACM, 2009: 810-814
[30] Stojanovic M, Freitag L. Multichannel Detection for Wideband Underwater Acoustic CDMA Communications [J]. IEEE Journal of Oceanic Engineering, 2006, 31(3): 685-695
[31] Proakis J G, Manolakis D K. Digital Signal Processing(4th Edition) [M]. NJ, USA: Prentice-Hall, 2006
[32] Cui Shu-guang, Madan R, Goldsmith A, et al. Joint Routing, MAC, and Link Layer Optimization in Sensor Networks with Energy Constraints[C]// Proc of IEEE ICC'05. New Jersey: IEEE, 2005: 725-729
[33] Tay Y C, Jamieson K, Balakrishnan H. Collision-Minimizing CSMA and Its Applications to Wireless Sensor Networks [J]. IEEE Journal on Selected Areas in Communications, 2004, 22(6): 1048-1057
[34] Rugin R, Mazzini G. A Simple and Efficient MAC-Routing Integrated Algorithm for Sensor Network[C]// Proc of IEEE Communications'04. New Jersey: IEEE, 2004: 3499-3503
[35] Skalli H, Ghosh S, Das S K, et al. Channel Assignment Strategies for Multiradio Wireless Mesh Networks: Issues and Solutions[J]. IEEE Communication Magazine, 2007, 45(11): 86-95
[36] Ma Jing, Zhang Ying-jun, Su Xin, et al. On capacity of wireless ad hoc networks with MIMO MMSE receivers [J]. IEEE Transactions on Wireless Communications, 2008, 7(12): 5493-5503
[37] Gong M X, Midkiff S F, Mao S. On-demand Routing and Channel Assignment in Multi-channel Mobile ad hoc Networks [J]. Ad Hoc Networks, 2009, 7(1): 63-78