基于概率的动态视图安全发布方法

宋金玲[1,2] 李芳玲[3] 刘国华[2,4] 黄立明[1] 张广斌[1] 王丹丽[2]
(河北科技师范学院 秦皇岛 066004)[1] (燕山大学信息科学与工程学院 秦皇岛 066004)[2]
(山东理工职业学院信息工程系 济宁 272017)[3] (东华大学计算机科学与技术学院 上海 201620)[4]

摘 要 视图发布的动态性和连续性使得视图间互相联系和影响。静态视图安全研究无法适应实际应用,如何保证动态视图的安全发布亟待解决。为了解决这个问题,首先提出了可能世界构造方法和隐私泄露概率计算方法,并给出了各种视图合并情况下的隐私泄露概率计算公式。然后,从相对安全的角度出发,给出了动态视图的安全判定公式。在此基础上,给出了动态视图的安全发布方法。所提方法能保证相对安全基础上的最大程度视图发布。

关键词 视图,安全,概率,隐私泄露
中图法分类号 TP309.2 **文献标识码** A

Security Dissemination Methods Based on Probability for Dynamic Views

SONG Jin-ling[1,2] LI Fang-ling[3] LIU Guo-hua[2,4] HUANG Li-ming[1] ZHANG Guang-bin[1] WANG Dan-li[2]
(Hebei Normal University of Science & Technology, Qinhuangdao 066004, China)[1]
(Department of Computer Science and Engineering, Yanshan University, Qinhuangdao 066004, China)[2]
(Department of Information Engineering, Shandong Polytechnic Vocational College, Jining 272017, China)[3]
(School of Computer Science and Technology, Donghua University, Shanghai 201620, China)[4]

Abstract Views can restrict database access on specifically attributes and tuples, so publishing views can reduce the possiblity of privacy disclosure. Howevere, the data owner will publish various views based on different application on different time, the dynamic and continuity of view dissemination made mutual contact and mutual influence among views. Considering the privacy disclosure of the publishing views themselves, which called static privacy disclosure, can not measure the disclosure accurately and can not adapt to the practical application, how to guarantee the security of the dynamic views dissemination must be resolved. To solve this problem, at first, the construction method of possible worlds and calculation method of privacy disclosure probability were propoded, and the calculating formulas of privacy disclosure probability under differrent conditions of views merging were presented. Then, the security determination formula for dynamic views was proposed from the point of relative sefe of view. Based on this, the safety dissemination method for dynamic views was presented, which can guarantee maximum level of views publishing upon relatively safe.

Keywords Views, Security, Probability, Privacy disclosure

1 引言

视图可以把用户对数据库的访问限制在特定的属性或元组上,这在一定程度上降低了隐私信息泄露的可能性[1]。但是,视图发布是一个动态的过程[2],数据持有者往往会根据不同的需求不断发布视图,即使发布的任意视图 V_i 相对隐私查询 S 都是安全的,攻击者通过多个视图的连接合并,仍有可能窥探到隐私信息,造成隐私信息的无意识泄露。因此,如何防止视图发布过程中的隐私泄露,保证发布视图的安全,成为数据库安全领域的一个重要课题。

视图发布过程中的安全问题包含静态视图安全和动态视图安全。静态视图安全指忽略已发布视图影响时发布视图集 V 本身的隐私泄露情况;动态视图安全指存在已发布视图集 V 时,发布视图集 V' 相对视图集 V 的隐私泄露情况。视图发布是一个动态、连续的过程,一方面已发布视图集 V 可作为先验知识对发布视图集 V' 的隐私泄露产生影响;另一方面,已发布视图集 V 的隐私泄露已经透明,发布视图集 V' 关注的应该是新的泄露。因此,静态视图安全研究无法准确反映发布视图的隐私泄露情况,不能适应实际应用。针对这种情况,本文利用概率思想,研究了动态视图的安全问题,提出了隐私泄露计算方法,并给出了动态视图安全判定公式,进而给出了动态视图的安全发布方法。所提方法能在相对安全的基础上保证最大程度的视图发布。

2 相关工作

目前,对视图安全的研究工作主要集中在视图的安全判定上,主要方法包括查询应答法[3]、基于概率[4]的安全判定方法和基于 K-匿名[5]的安全判定方法。

到稿日期:2010-10-25 返修日期:2011-02-25 本文受国家自然科学基金(60773100,61070032),河北省自然科学基金(F2009000475),河北科技师范学院科研创新团队建设经费(CXTD2010-05)资助。
宋金玲(1973-),女,博士生,讲师,主要研究领域为信息安全,E-mail:songjinling99@126.com;李芳玲(1976-),女,硕士,讲师,主要研究领域为信息安全;刘国华(1966-),男,博士,教授,博士生导师,CCF高级会员,主要研究领域为数据库理论、数据库安全、Web数据管理。

2001 年 Alon Halevy 提出用查询应答法[3]判断发布视图的安全性。查询应答法是将秘密查询 S 直接作用在视图组 V 上，如果能得到查询结果，就认为视图组 V 是不安全的。但是采用不可应答作为安全准则无法解决所有问题，有时虽然无法直接得到秘密查询的结果，但是通过视图间元组的连接、匹配等操作，攻击者依然会猜测出部分隐私信息。

在基于概率的安全判定方面，1977 年 Francois Bancilhon 等人首次利用条件概率[4]判定视图安全。2004 年，G. Miklau 和 D. Suciu 提出了用概率独立事件方法[6]判断待发布视图的安全性，首次从绝对安全角度来分析信息泄露，但是这种判定方法对安全性要求过于苛刻，无法适应实际应用。另外，该文的隐私泄露概率采用满足隐私查询的数据库实例个数和所有数据库实例个数的比值表示，数据库实例个数则根据属性域值的范围得到。此种数据库实例计算方法虽然在理论上是可行的，但是计算机计算能力有限，当属性个数和域值范围较大时，数据库实例的计算就无法实现。2005 年，NileshDalvi 等人用渐近条件概率方法[7]来判断发布视图安全，即利用 $\lim_{n\to\infty} P[S|V]$ 的值来判定发布视图的安全性。随着域值 D 大小 n 的逐渐增大，如果 $\lim_{n\to\infty} P[S|V]=0$，则认为发布视图安全；如果 $\lim_{n\to\infty} P[S|V]>0$，则认为信息泄露是不能忽略的。为了实现此判定算法，N. Dalvi 等人还详细分析了关联查询的概率计算，给出了在 $n\to\infty$ 时关联查询概率的算法，解决了用概率学知识判定发布视图安全性的关键问题。然而在 n 较小的情况下，这一方法的测量精度无法达到需求。同时，随着数据挖掘技术的应用，这种判定方法的缺陷更加明显。

在基于 K-匿名的安全判定方法方面，1998 年 Pierangela Samarati 和 L. Sweeney 首次提出了用 K-匿名保护模型方法[5]来判断发布视图安全。如果待发布视图满足 K-匿名约束，则认为视图是安全的，可以直接发布。2002 年，L. Sweeney 对该方法做了修改和补充[8]。同年，L. Sweeney 提出了 K-匿名保护模型无法满足的一些特例[9]，并提出了相应的解决方法。但该方法仅局限于判定单一视图发布的安全性，并没有考虑到同一个基本表上发布的所有视图之间的关系，用户有可能通过已发布的多个视图推测出超越其权限的数据；也没有考虑到数据的动态收集特性以及元组的动态添加、删除、改变和移动特征，一旦动态地添加、删除或改变元组，就需要对视图发布的安全性进行重新判定。2005 年，Yao Chao 等人将 K-匿名判定由单一视图扩展到多视图[10]，并分析了存在函数依赖情况下一些特例的 K-匿名判定，但并没有涉及到对信息泄露进行消除。2007 年，Li Ninghui 等人提出根据敏感信息的分布来判定视图是否存在信息泄露[11]。

现有视图安全研究都局限于静态视图的隐私泄露。为了更好地适应实际应用，本文研究了视图动态发布过程中的安全问题。采用视图间的元组匹配来构造可能世界，简化了隐私泄露概率计算。另外，视图安全判定不再苛求于绝对安全，保证了最大限度的视图发布。

3 基本符号和定义

A：一个属性。

Dom_A：属性 A 所有可能取的原子值（不可再分）构成的集合，称为属性 A 的域。

$|\mathrm{Dom}_A|$：属性 A 域的大小，即属性 A 不同取值的个数。

t：一个元组。

$t[A]$：属性 A 在元组 t 中的值。

$t[X]$：属性集 X 在元组 t 中的值。

$R(U)$：属性集为 U 的关系模式，R 为关系名。

$r(R)$：关系模式 $R(U)$ 的一个实例。

D：$D=\{R_1(U_1),R_2(U_2),\cdots,R_m(U_m)\}$ 为数据库模式。

d：$d=\{r_1(R_1),r_2(R_2),\cdots,r_m(R_m)\}$ 为数据库模式 D 的数据库实例。

$\sigma_\varphi(r(R))$：$\sigma_\varphi(r(R))=\{t\mid t\in r(R)$ 同时 t 满足条件 $\varphi\}$，σ 称为选择操作。

$\pi_X(r(R))$：$\pi_X(r(R))=\{t[X]\mid t\in r(R)\}$，$\pi$ 称为投影操作，本文的投影操作不包含重复元组。

$\bowtie(R_{i1},R_{i2},\cdots,R_{il}))$：$\bowtie(R_{i1},R_{i2},\cdots,R_{il}))=\{t(Y)\mid Y=\bigcup U_{ij}(1\leqslant j\leqslant l)$，且 $\exists t_j\in r_{ij}$ 使 $t(U_{ij})=t_j(U_{ij}),1\leqslant j\leqslant l\}$，$\bowtie$ 称为连接操作。

定义 1（视图） 对数据库模式 $D=\{R_1(U_1),R_2(U_2),\cdots,R_m(U_m)\}$ 的数据库实例 $d=\{r_1(R_1),r_2(R_2),\cdots,r_m(R_m)\}$，如果数据库模式 D 上的任一个 PSJ 查询可表示为 $E=\pi_X\sigma_\varphi\bowtie(R_{i1},R_{i2},\cdots,R_{il})$，则视图 V 表示查询 E 在数据库实例 d 上的执行结果，即物化视图。另外，E 也可表示为三元组 $E=(X,\varphi,R)$，其中 $X=\{A_1,A_2,\cdots,A_k\}$ 为投影属性集，φ 为选择条件，$R=\{R_{i1},R_{i2},\cdots,R_{il}\}$ 为连接的基本关系。

例 1 对表 1 所列的数据库实例 $d=\{r(Name,Sex,Age,Zip,Job,Problem)\}$，表 2－表 8 分别为数据库实例 d 的视图，其中各个视图的查询表达式分别为 $E_1=((Name,Zip,Job),Age<50,r)$，$E_2=((Problem),Age<50,r)$，$E_3=((Job,Problem),Age<50,r)$，$E_4=((Age,Problem),Age>40,r)$，$E_5=((Problem),Age>40,r)$，$E_6=((Job,Age,Problem),Age>40,r)$，$E_7=((Job,Problem),Age>40,r)$。

表 1 数据库实例 $d=\{r(Name,Sex,Age,Zip,Job,Problem)\}$

Name	Sex	Age	Zip	Job	Problem
Mary	Female	32	52000	Manager	AIDS
Jack	Male	23	18000	Doctor	Cold
Rose	Female	27	15000	Manager	Hepatitis
Mark	Male	35	22000	Foreman	Cancer
Nike	Male	42	28000	Doctor	Cancer
Bob	Male	44	23000	Doctor	Flu
Anne	Female	49	25000	Manager	HIV
LiLy	Female	52	52000	Nurse	Gastritis
John	Male	53	59000	Doctor	Cancer
Tom	Male	55	56000	Manager	Obesity

表 2 视图 V_1，其中 $E_1=((Name,Zip,Job),Age<50,r)$

TupleID	Name	Zip	Job
t_1	Mary	52000	Manager
t_2	Jack	18000	Doctor
t_3	Rose	15000	Manager
t_4	Mark	22000	Foreman
t_5	Nike	28000	Doctor
t_6	Bob	23000	Doctor
t_7	Anne	25000	Manager

表 3 视图 V_2，其中 $E_2=((Problem),Age<50,r)$

TupleID	Problem
t_1	AIDS
t_2	Cold
t_3	Hepatitis
t_4	Cancer
t_5	Flu
t_6	HIV

表 4　视图 V_3，其中 $E_3=((Job,Problem),\ Age<50,r)$

TupleID	Job	Problem
t_1	Manager	AIDS
t_2	Doctor	Cold
t_3	Manager	Hepatitis
t_4	Foreman	Cancer
t_5	Doctor	Cancer
t_6	Doctor	Flu
t_7	Manager	HIV

表 5　视图 V_4，其中 $E_4=((Age,Problem),\ Age>40,r)$

TupleID	Age	Problem
t_1	42	Cancer
t_2	44	Flu
t_3	49	HIV
t_4	52	Gastritis
t_5	53	Cancer
t_6	55	Obesity

表 6　视图 V_5，其中 $E_5=((Problem),\ Age>40,r)$

TupleID	Problem
t_1	Cancer
t_2	Flu
t_3	HIV
t_4	Gastritis
t_5	Obesity

表 7　视图 V_6，其中 $E_6=((Job,Age,Problem),\ Age>40,r)$

TupleID	Job	Age	Problem
t_1	Doctor	42	Cancer
t_2	Doctor	44	Flu
t_3	Manager	49	HIV
t_4	Nurse	52	Gastritis
t_5	Doctor	53	Cancer
t_6	Manager	55	Obesity

表 8　视图 V_7，其中 $E_7=((Job,Problem),\ Age>40,r)$

TupleID	Job	Problem
t_1	Doctor	Cancer
t_2	Doctor	Flu
t_3	Manager	HIV
t_4	Nurse	Gastritis
t_5	Manager	Obesity

定义 2(隐私查询，S)　隐私查询是指攻击者为探测隐私信息所进行的查询，用符号表示为 $S(IA:ia,SA:sa)$，其中 IA 是元组标识属性，SA 是敏感属性，ia 为具体标识值，sa 为具体敏感信息。$S(IA:ia,SA:sa)$ 表示探测 ia 的敏感信息是否是 sa，探测多组隐私信息用多个隐私查询表示。

例 2　对表 1 所列的数据库实例 d，隐私查询 $S(Name:$ Nike，$Problem:$Cancer)表示探测 Nike 是否患 Cancer 疾病。

注：针对隐私查询 $S(IA:ia,SA:sa)$，视图发布者的常识是属性 IA 和 SA 不能出现在同一视图中，以避免隐私的直接泄露。本文主要关注视图间直接连接导致的隐私泄露，为方便讨论，对下面任意视图 V 我们设 $IA\in V$ 或 $SA\in V$。

定义 3(分配，assign)　设视图 $V_1=\{t_1,t_2,t_3\}$，$V_2=\{t_1',t_2',t_3'\}$，则视图 V_1，V_2 元组间形成一一匹配(映射)关系，如 $\{(t_1:t_1'),(t_2:t_2'),(t_3:t_3')\}$，称为视图 V_1，V_2 的一个分配。

例 3　对 $V_1=\{t_1,t_2,t_3\}$，$V_2=\{t_1',t_2',t_3'\}$，可能的分配共有 6 种，分别为 $\{(t_1:t_1'),(t_2:t_2'),(t_3:t_3')\}$，$\{(t_1:t_1'),(t_2:t_3'),(t_3:t_2')\}$，$\{(t_1:t_2'),(t_2:t_1'),(t_3:t_3')\}$，$\{(t_1:t_2'),(t_2:t_3'),(t_3:t_1')\}$，$\{(t_1:t_3'),(t_2:t_1'),(t_3:t_2')\}$，$\{(t_1:t_3'),(t_2:t_2'),(t_3:t_1')\}$。

定义 4(可能世界，possible world)　设视图集为 $V_1,V_2,\cdots,V_n$，则对视图 $V_i,V_j\ (1\leqslant i,j\leqslant n)$ 按照某种分配进行连接合并后，将得到关于视图 V_i,V_j 的一个可能世界，记作 PW。

例 4　表 9 所列为视图 V_1(表 2)、V_3(表 4)所构成的一个可能世界，其中 V_1、V_3 中元组为顺序匹配。

表 9　可能世界示例

TupleID	Name	Zip	Job	Problem
t_1	Mary	52000	Manager	AIDS
t_2	Jack	18000	Doctor	Cold
t_3	Rose	15000	Manager	Hepatitis
t_4	Mark	22000	Foreman	Cancer
t_5	Nike	28000	Doctor	Cancer
t_6	Bob	23000	Doctor	Flu
t_7	Anne	25000	Manager	HIV

定义 5(隐私泄露概率)　给定隐私查询 $S(IA:ia,SA:sa)$ 和视图 V_i,V_j，如果 $ia\in V_i,sa\in V_j$，令 N_{PW} 表示视图 V_i,V_j 的所有可能世界个数，N_{PWS} 表示视图 V_i,V_j 满足隐私查询 S 的可能世界个数，则视图 V_i,V_j 相对隐私查询 S 的泄露概率为 $p(S|V_i,V_j)=\dfrac{|\text{满足隐私查询的可能世界}|}{|\text{可能世界}|}=\dfrac{N_{PWS}}{N_{PW}}$。

定义 6(多视图隐私泄露概率)　给定隐私查询 $S(IA:ia,SA:sa)$ 和视图集 $V=\{V_1,V_2,\cdots,V_n\}$，如果存在视图集 $(V_{i1},V_{j1})\subset V,(V_{i2},V_{j2})\subset V,\cdots,(V_{ik},V_{jk})\subset V$ 使 $ia\in V_{il},sa\in V_{jl}\ (1\leqslant l\leqslant k)$，即 $(V_{i1},V_{j1}),(V_{i2},V_{j2}),\cdots,(V_{ik},V_{jk})$ 均可能导致隐私泄露，则视图集 V 相对隐私查询 S 的泄露概率为 $p(S|V)=\underset{l=1}{\overset{k}{\mathrm{MAX}}}\ p(S|V_{il},V_{jl})$。

4　隐私泄露概率计算

设发布视图集为 $V=\{V_1,V_2,\cdots,V_n\}$，其中 $V_i=(X_i,\varphi_i,R_i)\ (1\leqslant i\leqslant n)$，隐私查询为 $S(IA:ia,SA:sa)$，如果存在 $V_i=(X_i,\varphi_i,R_i)$，$V_j=(X_j,\varphi_j,R_j)$ 使 $ia\in V_i,sa\in V_j$，则攻击者可通过 V_i,V_j 的元组分配构造可能世界，窥探隐私信息。按照视图 V_i,V_j 的查询表达式，V_i,V_j 隐私泄露概率的计算可分为查询条件相同、查询条件相交和不相交 3 种情况进行讨论。

4.1　选择条件相同($\varphi_i=\varphi_j$)

当 $\varphi_i=\varphi_j$ 时，视图 V_i 和 V_j 对应数据库中同一组元组，因此视图 V_i 和 V_j 中元组具有一一匹配关系。由于投影操作不包含重复元组，不失一般性，我们设 $|V_i|$ 和 $|V_j|$ 不同。

(1) 视图间无公共属性($X_i\cap X_j=\Phi$)

当 $X_i\cap X_j=\Phi$ 时，由于 V_i,V_j 没有公共属性(如视图 V_1 和 V_2)，因此攻击者在构造可能世界时，只能任意匹配 V_i 和 V_j 的元组。设 $N_{max}=\max(|V_i|,|V_j|)$，$N_{min}=\min(|V_i|,|V_j|)$，则 $N_{PW}=c_{N_{\max}}^{N_{\min}}N_{\min}!$。满足隐私查询的可能世界构造方式如下：$V_i$ 中 ia 所在元组与视图 V_j 中 sa 所在元组匹配，其它的元组任意匹配。因此，$N_{PWS}=c_{N_{\max}-1}^{N_{\min}-1}(N_{\min}-1)!$。根据 $P[S|\ V_i,V_j]=N_{PWS}/N_{PW}$，化简后有 $p(S|V_i,V_j)=1/N_{\max}$。

例 5　如表 2 和表 3 所列视图，视图 V_1 的查询表达式 $E_1=(\{Name,Zip,Job\},Age<50,r)$，视图 V_2 的查询表达式 $E_2=(\{Problem\},Age<50,r)$，$\varphi_1=\varphi_2$，$X_1\cap X_2=\phi$，因此 V_1,V_2 的选择条件相同且没有公共属性。设隐私查询为 $S=(Name:$Nike，$Problem:$ Cancer)，隐私泄露概率计算如下：

$N_{max}=\max(|V_1|,|V_2|)=7$，$N_{min}=\min(|V_1|,|V_2|)=6$，$p(S|V_1,V_2)=1/N_{max}=1/7$。

(2) 视图间存在公共属性($X_i\cap X_j\neq\Phi$)

当 $X_i\cap X_j\neq\Phi$ 时，视图 V_i 和 V_j 存在公共属性(如视图 V_1 和 V_3)，因此视图 V_i 和 V_j 元组可以按照公共属性进行等值匹配。由于隐私泄露只和隐私信息所在的公共属性分组有关，因此构造可能世界时只考虑该组元组的匹配即可，其他组元组的匹配可忽略。令 $X_i\cap X_j=A$，$t[IA]=ia$(即属性 IA 取值为 ia 的元组为 t)，如果 $t[A]=a$，设 $N_{max}=\max(|V_i(A:a)|,|V_j(A:a)|)$，$N_{min}=\min(|V_i(A:a)|,|V_j(A:a)|)$，则可求得 $N_{PW}=c_{N_{max}}^{N_{min}}N_{min}!$，$N_{PWS}=c_{N_{max}}^{N_{min}-1}(N_{min}-1)!$，故隐私泄露概率为 $p(S|V_i,V_j)=1/N_{max}$。

例 6　如表 2 和表 4 所列，视图 V_1 的查询表达式 $E_1=(\{Name,Zip,Job\},Age<50,r)$，视图 V_3 的查询表达式 $E_3=(\{Job,Problem\},Age<50,r)$，$\varphi_1=\varphi_3$，$X_1\cap X_3=\{Job\}$，因此 V_1，V_3 的选择条件相同且有公共属性。设隐私查询为 S(*Name*:Nike, *Problem*:Cancer)，隐私泄露概率计算如下：*name* 值为“Nike”的元组为 t_5，而 $t_5[Job]=$“Doctor”，由于 $N_{max}=N_{min}=|V_1(Job:\text{Doctor})|=|V_2(Job:\text{Doctor})|=3$，因此 $N_{PW}=3!$，$N_{PWS}==2!$，$P[S|V_1,V_3]=1/3$。

4.2　选择条件有交集($\varphi_i\cap\varphi_j\neq\Phi$ 且 $\varphi_i\neq\varphi_j$)

当视图 V_i 和 V_j 的选择条件有交集时，视图 V_i，V_j 中满足条件 $\varphi_i\cap\varphi_j$ 的元组对应数据库中同组元组，因此此组元组分配得到的可能世界才有意义。设 V_i'为 V_i 上满足条件 $\varphi_i\cap\varphi_j$ 的元组集，V_j'为 V_j 上满足条件 $\varphi_i\cap\varphi_j$ 的元组集，则只有当 $ia\in V_i'$，$sa\in V_j'$时，才可能产生隐私泄露。根据 V_i'，V_j'是否存在，隐私泄露概率的计算结果也会不同，下面对每种情况分别进行分析。

(1) 视图间无公共属性(即 $X_i\cap X_j=\Phi$)

当 V_i'，V_j'同时非空时，说明 V_i 和 V_j 均存在满足条件 $\varphi_i\cap\varphi_j$ 的元组，V_i 和 V_j 必同时包含 $\varphi_i\cap\varphi_j$ 中部分属性，这与 V_i 和 V_j 无公共属性矛盾。因此，当视图 V_i 和 V_j 不存在公共属性时，只有 V_i'或 V_j'非空和 V_i'，V_j'均空两种情况。

①$V_i'\neq\Phi$ 或 $V_j'\neq\Phi$

当 $V'_j\neq\Phi$ 且 $sa\in V_j'$时，即由视图 V_j 可以得到满足条件 $\varphi_i\cap\varphi_j$ 的元组，设 $|V_j'|=k$。此时，由于 V_i 无法得到满足条件 $\varphi_i\cap\varphi_j$ 的元组，V_i 中的任意 k 条元组都可能和 V_j'元组进行匹配构成可能世界。令 $|V_i|=N$，则有 $N_{PW}=c_N^k k!$，$N_{PWS}=c_{N-1}^{k-1}(k-1)!$，化简后 $p(S|V_i,V_j)=1/N$。同理，当 $V_i'\neq\Phi$ 且 $a\in V_i'$时，设 $|V_i'|=k$，$|V_j|=N$，则有 $N_{PW}=c_N^k k!$，$N_{PWS}=c_{N-1}^{k-1}(k-1)!$，化简后 $p(S|V_i,V_j)=1/N$。

例 7　如表 2 和表 5 所列视图，视图 V_1 的查询表达式 $E_1=(\{Name,Zip,Job\},Age<50,r)$，视图 V_4 的查询表达式 $E_4=(\{Problem\},Age>40,r)$，$\varphi_1\cap\varphi_4=\{40<Age<50\}\neq\phi$，$X_1\cap X_4=\phi$，因此 V_1，V_4 的选择条件有交集且没有公共属性。设隐私查询为 S=(*Name*:Nike, *Problem*:Cancer)，由于 V_4 中可以得到满足$\{40<Age<50\}$的元组，元组数目 $k=3$，另有 $N=|V_1|=7$，因此 V_1，V_4 的隐私泄露概率为 $p(S|V_1,V_4)=1/N=1/7$。

②$V_i'=\Phi$ 且 $V_j'=\Phi$

当视图 V_i 和 V_j 均得不到满足条件 $\varphi_i\cap\varphi_j$ 的元组时，V_i 和 V_j 用于匹配的元组数无法确定，可能为 $1,2,\cdots,\min(|V_i|,|V_j|)$，因此需要考虑不同匹配元组数下的隐私泄露概率。令 M_i 表示 V_i 和 V_j 匹配元组数为 k 的事件，则根据全概公式有 $p(S|V_i,V_j)=\sum_{k=1}^{\min(|V_i|,|V_j|)}p(M_k)p((S|V_i,V_j)|M_k)$。设事件 M_i 发生的概率相同，则 $p(M_k)=\frac{1}{\min(|V_i|,|V_j|)}$，代入上式有 $p(S|V_i,V_j)=\frac{1}{\min(|V_i|,|V_j|)}\sum_{k=1}^{\min(|V_i|,|V_j|)}p((S|V_i,V_j)|M_k)$。

当 V_i，V_j 匹配的元组数为 k 时，$N_{PW}=c_{|V_i|}^k c_{|V_j|}^k k!$，$N_{PWS}=c_{|V_i|}^{k-1}{}_{-1}c_{|V_j|-1}^{k-1}(k-1)!$，化简后 $p((S|V_i,V_j)|M_k)=\frac{N_{PWS}}{N_{PW}}=\frac{k}{|V_i||V_j|}$，$p(S|V_i,V_j)=\frac{1}{\min(|V_i|,|V_j|)}\sum_{k=1}^{\min(|V_i|,|V_j|)}\frac{k}{|V_i||V_j|}$。

例 8　如表 2 和表 6 所列视图，视图 V_1 的查询表达式 $E_1=(\{Name,Zip,Job\},\sigma_{Age<50},r)$，视图 V_5 的查询表达式 $E_5=(\{Problem\},Age>40,r)$，$\varphi_1\cap\varphi_5=\{40<Age<50\}\neq\phi$，$X_1\cap X_5=\phi$，因此 V_1，V_5 的选择条件有交集且没有公共属性。设隐私查询为 S=(*Name*:Nike, *Problem*:Cancer)，由于 V_1，V_5 中满足$\{40<Age<50\}$的元组都无法确定，且 $|V_1|=7$，$|V_5|=5$，因此 V_1 和 V_5 造成隐私泄露的概率为 $p(S|V_1,V_5)=1/\min(7,5)\sum_{k=1}^{\min(7,5)}k/|V_1||V_5|=1/5(1/7*5+2/7*5+3/7*5+4/7*5+5/7*5)=3/35$。

(2) 视图间存在公共属性($X_i\cap X_j\neq\Phi$)

当视图 V_i，V_j 存在公共属性时，元组间将按照公共属性进行等值匹配，因此隐私泄露概率的计算只和隐私信息所在的公共属性分组有关。

①$V_i'\neq\Phi$ 且 $V_j'\neq\Phi$

当 $V_i'\neq\Phi$，$V_j'\neq\Phi$ 且 $ia\in V_i'$，$sa\in V_j'$时，V_i，V_j 均能得到满足条件 $\varphi_i\cap\varphi_j$ 的元组，隐私泄露概率的计算只和 V'_i 和 V'_j 中隐私信息所在的公共属性分组有关。由于 V_i'和 V_j'的选择条件相同，隐私泄露概率公式与 4.1 节的第 2 种情况一致。令 $X_i\cap X_j=A$，$t[IA]=ia$(即属性 IA 取值为 ia 的元组为 t)，如果 $t[A]=a$，设 $N_{max}=\max(|V'_i(A:a)|,|V_j'(A:a)|)$，$N_{min}=\min(|V_i'(A:a)|,|V_j'(A:a)|)$，则隐私泄露概率为 $p(S|V_i,V_j)=1/N_{max}$。

②$V_i'\neq\Phi$ 或 $V_j'\neq\Phi$

当 $V'_j\neq\Phi$ 且 $sa\in V_j'$时，即视图 V_j 可以得到满足条件 $\varphi_i\cap\varphi_j$ 的元组，令 $X_i\cap X_j=A$，$t[SA]=sa$(即属性 SA 取值为 sa 的元组为 t)，如果 $t[A]=a$，设 $k=|V_j'(A:a)|$，$N=|V_i(A:a)|$，此时必有 $N\geqslant k$ 成立，则 $V_i(A:a)$中任意 k 条元组都可能和 $V_j'(A:a)$中元组进行匹配构成可能世界。可得到 $N_{PW}=c_N^k k!$，$N_{PWS}=c_{N-1}^{k-1}(k-1)!$，化简后 $p(S|V_i,V_j)=1/N$。同理，当 $V'_i\neq\Phi$ 且 $ia\in V_i'$时，令 $X_i\cap X_j=A$，$t[IA]=ia$(即属性 IA 取值为 ia 的元组为 t)，如果 $t[A]=a$，设 $k=|V_i'(A:a)|$，$N=|V_j(A:a)|$，则 $p(S|V_i,V_j)=1/N$。

例 9　如表 2 和表 7 所列视图，视图 V_1 的查询表达式 $E_1=(\{Name,Zip,Job\},\sigma_{Age<50},r)$，视图 V_6 的查询表达式 $E_6=(\{Job,Age,Problem\},Age>40,r)$，$\varphi_1\cap\varphi_6=\{40<Age<50\}\neq\phi$，$X_1\cap X_6=Job$，因此 V_1，V_6 的选择条件有交集且存在公共属性。设隐私查询为 S=(*Name*:Anne, *Problem*:HIV)，由于 V_6 中满足$\{40<Age<50\}$的元组有 3 个，且 *Problem* 为 HIV

的元组中 Job 的取值为 Manager，在视图 V_6 中 Job 值为 $Manager$ 的元组数目为 3，因此 V_1，V_6 的隐私泄露概率为 $p(S|V_1,V_6)=1/N=1/3$。

③$V_i'=\Phi$ 且 $V_j'=\Phi$

当无法确定视图 V_i 和 V_j 满足条件 $\varphi_i \cap \varphi_j$ 的元组集时，视图 V_i，V_j 只能按隐私信息所在的公共属性分组进行等值匹配。令 $X_i \cap X_j=A$，$t[IA]=ia$（即属性 IA 取值为 ia 的元组为 t），如果 $t[A]=a$，设 $V_i''=V_i(A:a)$，$V_j''=V_j(A:a)$，则 V_i 和 V_j 间匹配的元组集为 V_i''和 V_j''，且匹配的元组数可能为 $1,2,\cdots,\min(|V_i''|,|V_j''|)$，令 M_k 表示 V_i 和 V_j 匹配元组数为 k 的事件，且 M_k 发生的概率相同，则根据全概公式，$p(S|V_i,V_j)=1/\min(|V_i''|,|V_j''|)\sum_{k=1}^{\min(|V_i''|,|V_j''|)} p((S|V_i,V_j)|M_k)$。当 V_i，V_j 匹配元组数为 k 时，$N_{PW}=c_{|V_i''|}^{k}c_{|V_j''|}^{k}k!$，$N_{PWS}=c_{|V_i''|-1}^{k-1}c_{|V_j''|-1}^{k-1}(k-1)!$，$p((S|V_i,V_j)|M_k)=N_{PWS}/N_{PW}=k/|V_i''||V_j''|$，代入得 $p(S|V_i,V_j)=1/\min(|V_i''|,|V_j''|)\sum_{k=1}^{\min(|V_i''|,|V_j''|)} k/|V_i''||V_j''|$。

例 10　如表 2 和表 8 所列视图，视图 V_1 的查询表达式 $E_1=(\{Name,Zip,Job\},\sigma_{Age<50},r)$，视图 V_7 的查询表达式 $E_7=(\{Job,Problem\},Age>40,r)$，$\varphi_1 \cap \varphi_7=\{40<Age<50\}\neq\phi$，$X_1 \cap X_7=Job$，因此 V_1，V_7 的选择条件有交集且存在公共属性。设隐私查询为 $S=(Name:\text{Anne},Problem:\text{HIV})$，由于 V_1，V_7 中满足 $\{40<Age<50\}$ 的元组都无法确定，V_1 中 $Name$ 值为 Anne 的元组中 Job 取值为 Manager，$|V_1''|=|V_1(Job:\text{Manager})|=3$，$|V_7''|=|V_7(Job:\text{Manager})|=2$，因此 V_1，V_7 的隐私泄露概率为 $p(S|V_1,V_7)=1/\min(3,2)\sum_{k=1}^{\min(3,2)} k/|V_1''||V_7''|=1/2(1/3*2+2/3*2)=1/4$。

4.3　选择条件无交集（$\varphi_i \cap \varphi_j=\Phi$）

当视图 V_i 和 V_j 的选择条件没有交集时，视图 V_i 和 V_j 的选择条件必定相悖（如 $Age>40$ 和 $Age<30$），说明视图 V_i 和 V_j 取自数据库实例 d 的不同元组。此时，不论是否存在公共属性，V_i 和 V_j 根据元组分配构造的可能世界都没有实际意义，因此视图不存在隐私泄露。

5　动态视图安全发布

5.1　动态视图安全发布方法

根据条件概率的思想，本节给出了动态视图的安全发布方法。

在已发布视图集 $V=\{V_1,V_2,\cdots,V_n\}$ 的基础上，再发布视图集 $V'=\{V_1',V_2',\cdots,V_m'\}$（其中最多存在一对视图 V_k'，$V_l'\in V'$ 使 $ia\in V_k'$，$sa\in V_l'$）。如果 V' 对推测隐私信息没有新的贡献，即视图集 V' 发布前后得到隐私信息的概率没有变化，则可认定发布视图是安全的。

定理 1　假定 P 为隐私泄露概率，S 是隐私查询，$V=\{V_1,V_2,\cdots,V_n\}$ 是已发布视图集，$V'=\{V_1',V_2',\cdots,V_m'\}$ 为待发布视图集，若有：

$$P[S|V]=P[S|V\cup V'] \tag{1}$$

则在已发布视图集 V 条件下，发布视图集 $V'=\{V_1',V_2',\cdots,V_m'\}$ 关于隐私查询 S 是安全的。

证明：如果 $P[S|V]=P[S|V\cup V']$，则说明隐私查询 S 在视图集 V 和视图集 $V\cup V'$ 上的泄露概率相同，即发布视图 V' 后，视图集 $V\cup V'$ 并没有比视图集 V 泄露更多的关于秘密查询 S 的信息。因此，如果上述条件成立，可知在已发布视图集 V 条件下，发布视图 V' 关于隐私查询 S 是安全的。定理 1 得证。

例 11　根据 4.2 节的示例可知，如果首先发布视图集 V_1，V_3，则有 $P[S|V_1,V_3]=1/3$。再发布视图 V_2 时，由于 $P[S|V_1,V_2,V_3]=\max(P[S|V_1,V_2],P[S|V_1,V_3])=\max(1/7,1/3)=1/3$，此时有 $P[S|V_1,V_3]=P[S|V_1,V_2,V_3]$，则在先发布视图 V_1，V_3 的条件下，发布视图 V_2 是安全的。同理，如果首先发布视图集 V_1，V_2，则有 $P[S|V_1,V_2]=1/7$。再发布视图 V_3 时，由于 $P[S|V_1,V_2,V_3]=1/3$，此时有 $P[S|V_1,V_2,V_3]>P[S|V_1,V_3]$，则在先发布视图 V_1，V_2 的条件下，发布视图 V_3 是不安全的。

由于视图集 V 作为先验知识的影响，因此式(1)并不一定总是成立的，即发布视图集 V' 可能存在一定的隐私泄露。但是，在实际应用中，相对较小的隐私泄露用户是可以接受的。因此，为了最大程度地发布视图，有必要先测量视图集 V' 的隐私泄露，再发布泄露在用户允许范围内的视图。根据定理 1，我们采用下式测量视图集 V' 的隐私泄露程度：

$$Leak(S,V')=\begin{cases}P[S|V\cup V']-P[S|V], & P[S|V\cup V']>P[S|V]\\ 1, & P[S|V\cup V']=1\end{cases} \tag{2}$$

根据视图安全判定定理和隐私泄露测量公式，本文从相对安全角度出发，采用下面的视图安全发布方法：设用户可接受的隐私泄露阈值为 θ。当 $Leak(S|V')\leqslant\theta$ 时，就认为视图集 V' 在视图集 V 下是安全的，可发布视图集 V'；当 $Leak(S|V')>\theta$ 时，则认为视图集 V' 在视图集 V 下是不安全的，需先删除 V' 中的 ia 和 sa，再进行发布。

上述视图安全发布方法是从相对安全角度出发的，因此它可以最大限度地发布视图，适合实际应用。

5.2　视图安全发布算法 PSV

根据上面给出的视图安全发布方法，下面将给出具体的视图安全发布算法。首先根据 4.2 节的隐私泄露概率公式，给出 $P[S|V_i,V_j]$ 的求解算法 PD（见算法 1）。

算法 1　隐私泄露概率求解算法 PD。

```
PRIVACY DISCLOURE(S,Vi,Vj)
输入:隐私查询 S(IA:ia,SA:sa),视图 Vi=(Xi,φi,Ri),Vj=(Xj,φj,
     Rj);
输出:P[S|Vi,Vj]
Step1  初始化 Nmax=0;N=0;P[S|Vi,Vj]=0;A=Φ;
Step2
  /* 视图 Vi,Vj 选择条件相同时的隐私泄露概率计算 */
  if φi=φj then
     {if Xi∩Xj=Φ then /* 视图 Vi,Vj 不存在公共属性 */
     { Nmax=max(|Vi|,|Vj|);
     P(S|Vi,Vj)=1/Nmax;}
  else /* 视图 Vi,Vj 存在公共属性 */
     { A=Xi∩Xj;
       t={ t'| t'[IA]=ia};
       a= t[A];
       Nmax=max(|Vi(A:a)|,|Vj(A:a)|);
       P(S|Vi,Vj)=1/Nmax;}
       }
  /* 视图 Vi,Vj 选择条件相交时的隐私泄露概率计算 */
```

else if $\varphi_i \cap \varphi_j \neq \Phi$

{V_i'={V_i 上满足条件 $\varphi_i \cap \varphi_j$ 的元组集};

V_j'={V_j 上满足条件 $\varphi_i \cap \varphi_j$ 的元组集};

if $X_i \cap X_j = \Phi$ then /* 视图 V_i,V_j 不存在公共属性 */

{/* V_j 可确定满足条件 $\varphi_i \cap \varphi_j$ 元组集 */

if $V_j' \neq \Phi$ and sa$\in V_j'$ then N=$|V_i|$;

/* V_i 可确定满足条件 $\varphi_i \cap \varphi_j$ 元组集 */

if $V_i' \neq \Phi$ and ia$\in V_i'$then N=$|V_j|$;

$P(S|V_i,V_j)=1/N$;

/* 视图 V_i,V_j 均无法确定满足条件 $\varphi_i \cap \varphi_j$ 元组集 */

if $V_i' = \Phi$ and $V_j' = \Phi$ then

$$P(S|V_i,V_j)=1/\min(|V_i|,|V_j|)\sum_{k=1}^{\min(|V_i|,|V_j|)} k/|V_i||V_j|;\}$$

else /* 视图 V_i,V_j 存在公共属性 */

{ $A=X_i \cap X_j$;

/* 视图 V_i,V_j 均可确定满足条件 $\varphi_i \cap \varphi_j$ 元组集 */

if $V_i' \neq \Phi$ and $V_j' \neq \Phi$ and ia$\in V_i'$ and sa$\in V_j'$ then

{ t={t'|t'$\in V_i'$,t'[IA]=ia};

a= t[A];

N_{max}=max(|V_i'(A;a)|,| V_j'(A;a)|);

$P(S|V_i,V_j)=1/N_{max}$;}

/* 视图 V_j 可确定满足条件 $\varphi_i \cap \varphi_j$ 元组集 */

if $V_j' \neq \Phi$ and sa$\in V_j'$ then

{ t={t'|t'$\in V_j'$,t'[SA]=sa};

a=t[A];

N=| V_i(A;a) |;

$P(S|V_i,V_j)=1/N$;}

/* 视图 V_i 可确定满足条件 $\varphi_i \cap \varphi_j$ 元组集 */

if $V_i' \neq \Phi$ and ia$\in V_i'$ then

{ t={t'|t'$\in V'_i$,t'[IA]=ia};

a=t[A];

N=| V_j(A;a) |;

$P(S|V_i,V_j)=\frac{1}{N}$;}

/* 视图 V_i,V_j 均无法确定满足条件 $\varphi_i \cap \varphi_j$ 元组集 */

if $V_i' = \Phi$ and $V_j' = \Phi$ then

{ t={t'|t'$\in V_i$,t'[IA]=ia};

a=t[A];

$V_i''=V_i(A;a)$;

$V_j''=V_j(A;a)$;

$$P(S|V_i,V_j)=1/\min(|V_i''|,|V_j''|)\sum_{k=1}^{\min(|V_i''|,|V_j''|)} k/|V_i''||V_j''|;\}$$

}}

else /* 视图 V_i,V_j 选择条件不相交时隐私泄露概率 */

$P[S|V_i,V_j]=0$;

Step3 return($P[S|V_i,V_j]$)。

使用上述算法,可以得到任意两个视图的隐私泄露概率,通过分别计算视图集 V 和 $V \cup V'$ 中的隐私泄露概率,并用 Leak 公式测量视图集 V'的隐私泄露强度。当 Leak 值不超出设定值 θ 时,视图集 V'相对安全,可以发布,否则禁止发布。下面给出视图安全发布算法 PSV(见算法 2)。

算法 2 视图安全发布算法 PSV。

PUBLISHING SECURE VIEW(S,V,V',θ)

输入:隐私查询 S(IA;ia,SA;sa),已发布视图集 V,待发布视图集 V',隐私泄露阈值 θ;

输出:相对安全的视图集 V'。

Step1 初始化 p=0;p_{max1}=0;p_{max2}=0;

Step2 /* 视图集 V 的隐私泄露概率计算 */

for each V_i,$V_j \in V$,do

{if ia$\in V_i$ and sa$\in V_j$ then p= PRIVACY DISCLOURE(S,V_i,V_j);

if p>p_{max1} then p_{max1}=p;};

Step3 /* 视图集 V$\cup$V'的隐私泄露概率计算 */

$p_{max2}=p_{max1}$;

V_k={ V_i| $V_i \in V'$,ia$\in V_i$};

V_l={ V_j| $V_j \in V'$,sa$\in V_j$};

if $V_k \neq \Phi$ and $V_l \neq \Phi$ then p=PRIVACY DISCLOURE(S,V_k,V_l);

if p>p_{max2} then p_{max2}=p;

for each V_i where $V_i \in V$ and ia$\in V_i$,do

{ p= PRIVACY DISCLOURE(S,V_i,V_l);

if p>p_{max2} then p_{max2}=p;}

for each V_j where $V_j \in V$ and sa$\in V_j$,do

{ p= PRIVACY DISCLOURE(S,V_k,V_j);

if p>p_{max2} then p_{max2}=p;};

Step4 Leak= $p_{max2}-p_{max1}$;/* Leak 值计算 */

if Leak<=θ then 发布视图集 V';

else

{删除 V'中的 ia 和 sa;

发布视图集 V';}。

5.3 算法实例

例 12 设视图集 V_1,V_2 为已发布视图,隐私查询为 S=(*Name*:Nike,*Problem*:Cancer),隐私泄露阈值 θ=0.1,如果需要再发布视图 V_3,则 PSV 算法的执行流程如下:首先执行第 2 步,得到 $p_{max1}=P[S|V_1,V_2]=1/7$;然后执行第 3 步,计算得到 $p_{max2}=P[S|V_1,V_3]=1/3$;其次执行第 4 步,得到 $Leak=p_{max2}-p_{max1}=0.19$。由于 $Leak<=\theta$ 不成立,因此认为视图 V_3 是不安全的,禁止发布视图 V_3。

例 13 设视图集 V_1,V_5 为已发布视图,隐私查询为 S=(*Name*:Anne,*Problem*:HIV),隐私泄露阈值 θ=0.1,如果需要再发布视图 V_4,则 PSV 算法的执行流程如下:首先执行 Step2,得到 $p_{max1}=P[S|V_1,V_5]=3/35$;然后执行 Step3,计算得到 $p_{max2}=P[S|V_1,V_3]=1/7$;其次执行 Step4,得到 $Leak=p_{max2}-p_{max1}=0.057$。由于 $Leak<=\theta$ 成立,因此认为视图 V_4 是安全的,发布视图 V_4。

5.4 算法优化

根据 PSV 算法描述,PSV 算法时间复杂度主要由 Step2 和 Step3 决定。在 Step2 中,当包含 *ia* 和 *sa* 的视图各占 $n/2$ 时循环次数最多,即每次只发布一对包含 *ia* 和 *sa* 的视图。由于 PRIVACY DISCLOURE 算法中均为分支语句,其时间复杂度为常数,记为 O(1),Step2 时间复杂度最坏为 $n/2*n/2*O(1)$,即 $O(n^2)$。Step3 时间复杂度最坏为 $n/2*O(1)+n/2*O(1)$,即 $O(n)$。因此,PSV 算法的时间复杂度为 $O(n^2)$。

在上述 PSV 算法中,Step2 用于计算已发布视图集的隐私泄露概率。但是对当前调用的 PSV 算法而言,前次调用 PSV 算法时发布视图 V'后的隐私泄露概率即为已发布视图隐私泄露概率。因此,如果在每次调用 PSV 算法时将其保存下来,Step2 可省略。此时 PSV 算法的时间复杂度变为 $O(n)$。改进后的 PSV 算法描述如下:

(下转第 167 页)

el Interchange)的格式导出。此 XMI 文件可以被理解的第三方分析工具导入并进行进一步的多维展示和操作。图 4 给出模型构建过程。

图 5 给出了一个转化规则的示例片段。这段代码主要是从定义链接库中按照 DefinitionLinkArc 的描述，整理概念之间的父子层次关系，形成维度及其层次结构，这是构成图 3 中构造型〈〈StructureMap〉〉的一部分。

结束语 本文给出了一种基于 CWM 的财务数据多维分析模型的构建方法，通过此模型可以将缺少必要语义描述的 XBRL 财务数据报告进行转化和集成，生成由商业智能标准描述的多维模型。这多维模型可以被工业界成熟的其它数据分析工具，如数据仓库、OLAP 分析等所识别并进一步进行数据的多维展现和分析，使得基于 XBRL 商业报告的多维分析更为方便。本系统采用 XQuery 描述模型的构建规则并统一管理，因此具有较好的通用性及开放性。

参 考 文 献

[1] XBRL 国际组织. Extensible Business Reporting Language (XBRL) 2.1[S]. 2008

[2] Poole J, Mellor D, Chang D, et al. Common Warehouse Metamodel[M]. Wiley, John & Sons, 2001

[3] Lara R, Cantador I, Castells P. XBRL taxonomies and OWL ontologies for investment funds[C]// Tucson A Z. ed. 25th International Conference on Conceptual Modeling(ER 2006). Springer Verlag, 2006

[4] Spies M. An ontology modelling perspective on business reporting[J]. Information Systems, 2010, 35(4): 404-416

[5] OMG. Common Warehouse Metamodel Specification [S]. 2003

[6] 赵晓非，黄志球，沈国华，等. 基于元模型的工程数据仓库系统元数据集成[J]. 南京航空航天大学学报，2006，38(3)：341-346

[7] 田翀，宁洪，李姗姗，等. CWM 研究及相关元数据管理系统的设计[J]. 计算机工程，2006，32(11)：100-102

[8] 夏秀峰，孙娜，石祥滨，等. 基于 CWM 的结构化异构数据抽取方法研究与实现[J]. 计算机应用与软件，2009，26(12)：108-110

[9] 陈兴建，郝文宁，靳大尉，等. 基于 CWM 和 EMF 的数据库元数据处理[J]. 计算机工程，2010，36(13)：40-41

[10] Liu Z H, Baby T, Krishnamurthy S, et al. XBRL repository-An industrial approach of management of XBRL documents[C]// 26th IEEE International Conference on Data Engineering(ICDE 2010). IEEE Computer Society, 2010: 1037-1047

[11] Ma J, Pan X, Yao J. The research on the knowledge representation model of XBRL[C]// IEEE International Conference on Service Operations and Logistics, and Informatics(IEEE/SOLI 2008). IEEE Computer Society, 2008

(上接第 163 页)

PUBLISHING SECURE VIEW(S, p_V, V′, θ)

输入：隐私查询 S(IA: ia, SA: sa)，已发布视图集 V 的隐私泄露概率 p_V，待发布视图集 V′，隐私泄露阈值 θ；

输出：相对安全的视图集 V′。

Step1 初始化 $p=0$；$p_{max}=p_V$；

Step2 /* 视图集 V∪V′的隐私泄露概率计算 */

$V_k=\{ V_i \mid V_i\in V', ia\in V_i\}$；

$V_l=\{ V_j \mid V_j\in V', sa\in V_j\}$；

if $V_k\neq\Phi$ and $V_l\neq\Phi$ then $p=$ PRIVACY DISCLOURE(S, V_k, V_l)；

if $p>p_{max}$ then $p_{max}=p$；

for each V_i where $V_i\in V$ and $ia\in V_i$, do

{ $p=$ PRIVACY DISCLOURE(S, V_i, V_l)；

if $p>p_{max}$ then $p_{max}=p$；}

for each V_j where $V_j\in V$ and $sa\in V_j$, do

{ $p=$ PRIVACY DISCLOURE(S, V_k, V_j)；

if $p>p_{max}$ then $p_{max}=p$；}；

Step3 $Leak=p_{max}-p_V$；/* Leak 值计算 */

if $Leak<=\theta$ then

{发布视图集 V′；

$p_V=p_{max}$；}

else

{删除 V′中的 ia 和 sa；

发布视图集 V′；}；

Step4 return(p_V)。

结束语 本文从相对安全角度出发，利用条件概率思想，综合考虑了视图间的联系和制约关系，研究了动态视图的隐私泄露问题。本文的视图由一般的 PSJ 查询表达式生成，具有较好的实用性。提出的可能世界构造方法使隐私泄露概率计算更容易实现，也更具有实际意义。后期工作将探索隐私查询其他表示时的隐私泄露概率计算公式，使该公式更具一般意义。

参 考 文 献

[1] Sweeney L. K-Anonymity: A model for protecting privacy [J]. In'l Journal on Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 557-570

[2] Dalenius T, Reiss S. Data swapping: A technique for di- sclosure control [J]. Journal of Statistical Planning and Inference, 1982, 6(1): 73-85

[3] Halevy A. Answering Queries Using Views: A survey [J]. VLDB Journal, 2001, 10(4): 270-294

[4] Bancilhon F, Spyrators N. Protection of Information in Relational Data Bases[C]// VLDB. Tokyo. Japan, 1977: 494-500

[5] Samarati P, Sweeney L. Protecting Privacy When Disclosing Information: k-anonymity and Its Enforcement through Generalization and Suppression[R]. SRL-CSL-98-04. SRI Computer Science Laboratory, 1998

[6] Miklau G, Suciu D. A Formal Analysis of Information Disclosure in Data Exchange[C]// Proceedings of the 20th ACM SIGMOD International Conference on Management of Data. Orlando, USA, 2004: 507-534

[7] Dalvi N, Miklau G, Suciu D. Asymptotic Conditional Probabilities for Conjunctive Queries [C]// Proceedings of the Sixth International Conference on Database Theory. Edinburgh, UK, 2002: 289-305

[8] Sweeney L. Achieving k-anonymity privacy protection using gen-eralization and suppression [J]. Uncertainty, Fuzziness and Knowledge-based Systems, 2002, 10(5): 571-578

[9] Sweeney L. K-anonymity: A Model for Protecting Privacy[J]. Int'l Journal on Uncertainty, Fuzziness, and Knowledgebased Systems, 2002, 10(5): 557-570

[10] Yao Chao, Wang Xiao-yang Sean, et al. Checking for K-Anonymity Violation by Views[C]// VLDB. 2005: 910-921

[11] Li Ning-hui, Li Tian-cheng, Venkatasubramanian S. t-Closeness: Privacy Beyond k-Anonymity and l-diversity[C]// Proceedings of the 23th International Conference on Data Engineering. Istanbul, Turkey, 2007: 106-115