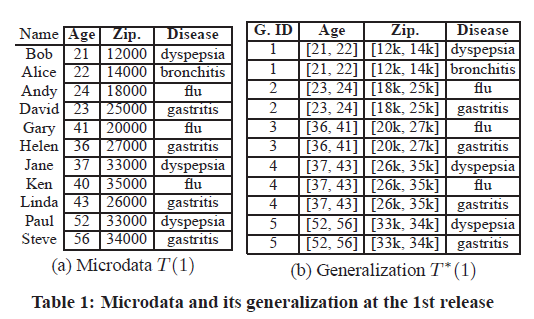
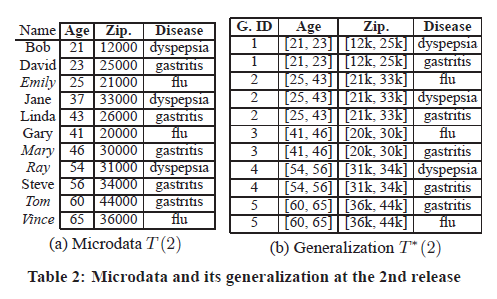
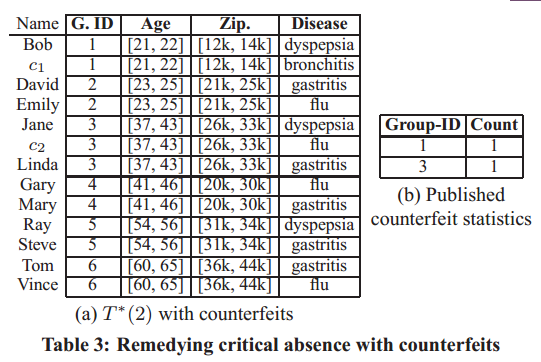
Table1b是2-diverse的





整合了m-不变式和伪造泛化，后者技巧促进前者的执行。



不同时期都存在的元组，其所在不同发布版本表的等价组中，具有相同的敏感值域。（不变性）

保持上述不变性。。。

工作脉络

1.形式化概念

2.分析隐私保护理论，说明攻击者可以如何攻击，证明m不变性具有优势

3.设计遵循m-不变式的算法，并通过1.最小化伪造元组数2.等价组包含的个体数来

最大化信息利用率。

算法是增量的，只依赖上一次，无需保留在上次之前的发布

1.形式化

Aid Aqi As

T(j)

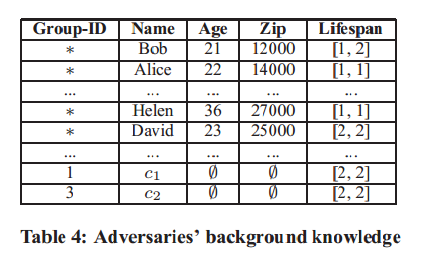
{T ∗(j), R(j)}

Interval在分类型属性上的定义：将分类类型转化为数值类型，进行值的排序。

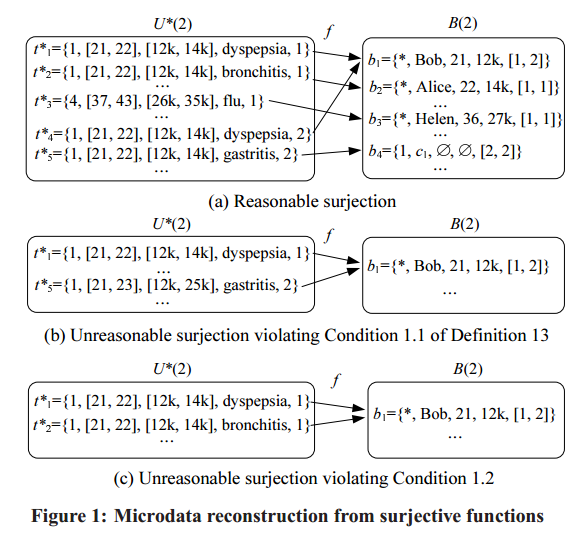
定义5：U(n)：用于表示之前发布的匿名集合的并集，故在不同T(j)中出现的相同t，只在U(n)中出现一次。其中的每一个t，都有一个生存周期[x,y];

定义6：先验知识 n时刻包含：

1. 使用的泛化原则
2. 知识表B（n）由 Ag 、Al（生存周期） 、 U(n)中除As组成，同时包含伪造元组的细节。



定义8：重发布：根据之前的T\*(j)和R(j)（1<=j<n）来计算T\*(n)和R(n)。



DEFINITION 13 (REASONABLE SURJECTION). A function f

in Definition 10 is reasonable, if it satisfies these conditions:

1. the following holds for each row b ∈ B(n): Given the set

f−1(b) of tuples t∗ ∈ U∗(n) satisfying f(t∗) = b, then

1.1. all tuples in f−1(b) carry the same sensitive value;

1.2. for any timestamp j in the lifespan b[Al] of b, there

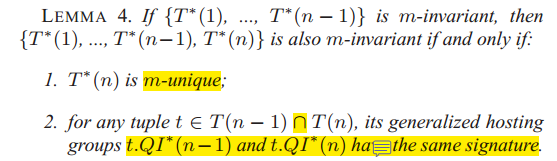
exists a unique tuple t∗ ∈ f−1(b) with t∗[Atm] = j.

1. f−1 decides a possible generalization instance that conforms to the deployed generalization principle (Definition 4).

定义15：t.CSS(j) t在j时刻的候选敏感集合指：T\*(j)中的所有QI\*中敏感值的集合。

引理1：只有一个单独的元素，则其泄露概率为1；

LEMMA 1. *Let* t *be a tuple in* U(n) *with a lifespan* [x, y]*. Regardless of the generalization principle applied,* risk(t) = 1*, if  
only a single element exists in* t.CSS(x) ∩ t.CSS(x + 1) ∩ ... ∩  
t.CSS(y)*.*

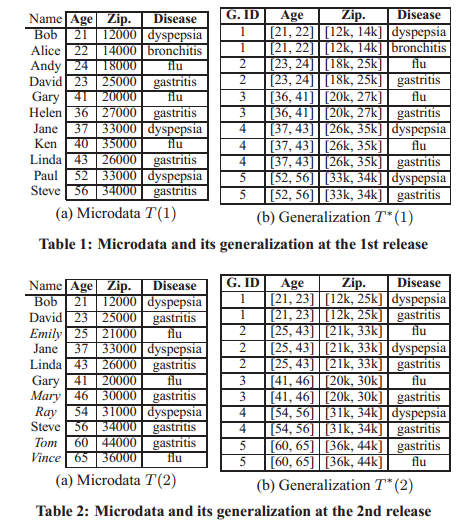


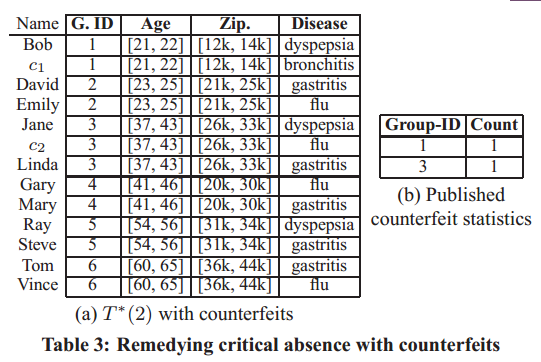
m-unique 表示 至少含有m个不同的敏感值

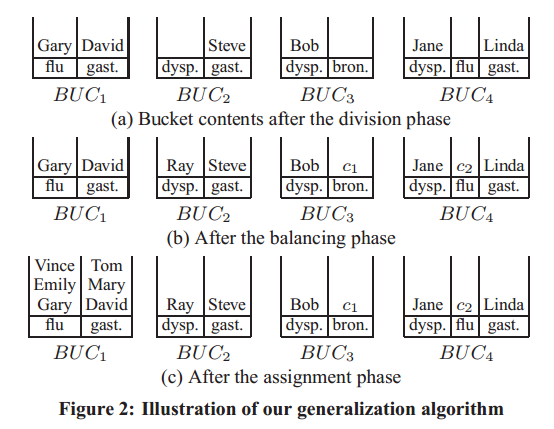
signature 表示 QI\*组的敏感值的set集合 ；t的signature为T\*(n-1)中其所在QI\*组的signature

1. 算法实现（m-不变式）

Intuitive 直观的目标：1.最少数量的伪造元组2.最少的泛化来使QI值失真最少







Division，把T(n)和T(n-1)交集中signature相同的t，分到同一个桶

Balance：在一个桶里 每个敏感值都 被相同数量的元组所拥有

重点：分配阶段接受前一个阶段（balance）的s\_作为参数，

一次将a\*b 个元组从s\_中移到 含有至少b》=m个敏感值的桶中。

Initially, we set it to the  
smallest possible value m, and attempt to solve α from the above  
three inequalities. If α exists, then the formulation of α and β  
is completed. Otherwise, β is increased by 1, and we again try to  
solve α. This process is repeated, until β reaches the first value that  
yields a solution of α.