# HW1

PB18111760 王嘉梁

## 1.(10’) Try to explain why recursive (c, l)-diversity guards against all adversaries who possess at most l − 2 statements of the form “Bob does not have heart disease”.

设等价类中的敏感属性有$m$种取值$S_{1...m}$, $r_i$表示出现次数第$i$多的取值的频次

对于$(c,l)-diversity$，有$r_1 < c(r_l + r_{l+1} + ... + r_m)$

若攻击者知晓$l-2$条有关知识，假设可以推理出不为$S_2...S_{l-1}$

则$r1 + r_l + ... + r_m = 1$，由于$r_1 < c(r_l + r_{l+1} + ... + r_m)$

等价类中剩余敏感值出现频次不会有悬殊差距，故攻击者无法获得额外信息。

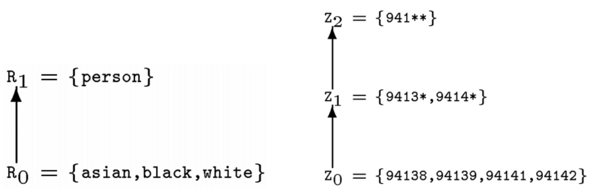
若攻击者掌握$l-1$条有关知识，可以推理出不为$S_2...S_l$

$r1 + r_{l+1} + ... + r_m = 1$

但由已知等式，不能保证$r_1$和$(r_{l+1} + ... + r_m)$相差小于$c$倍

所以攻击者获得了额外的信息。

**2.(15’) Consider domains R0 (Race) and Z0 (ZIP code) whose generalization hierarchies are illustrated in Fig. 1a and Fig. 1b independently. Assume QI = {Race, ZIP} to be a quasi-identifier. Consider private table P T illustrated in table 1, please give all possible 2-anonymity using full domain generalization and suppression under the condition that the maximum number of suppressed records (MaxSup) is less than or equal to 1. (If it is not generalized, 4 records need to be suppressed, which does not meet the requirement of MaxSup ≤ 1, illustrated in table 2).**




(a) The generalization hierarchy of $R_0$ (b) The generalization hierarchy of $Z_0$

Figure 1: Generalization hierarchies



| Race:$R_0$ | ZIP:$Z_0$ |
|---|---|
| asian | 94138 |
| asian | 94138 |
| asian | 94142 |
| asian | 94142 |
| black | 94138 |
| black | 94141 |
| black | 94142 |
| white | 94138 |

Table 1: $PT$

| Race:$R_0$ | ZIP:$Z_0$ |
|---|---|
| asian | 94138 |
| asian | 94138 |
| asian | 94142 |
| asian | 94142 |
| | |
| | |
| | |
| | |

Table 2: Suppression for table $PT$

| Race:R0 | ZIP:Z0 |
| --- | --- |
| person | 94138 |
| person | 94138 |
| person | 94142 |
| person | 94142 |
| person | 94138 |
| suppression | suppression |
| person | 94142 |
| person | 94138 |

| Race:R0 | ZIP:Z0 |
| --- | --- |
| asian | 941** |
| asian | 941** |
| asian | 941** |
| asian | 941** |
| black | 941** |
| black | 941** |
| black | 941** |
| suppression | suppression |

| Race:R0 | ZIP:Z0 |
| --- | --- |
| person | 9413* |
| person | 9413* |
| person | 9414* |
| person | 9414* |
| person | 9413* |
| person | 9414* |
| person | 9414* |
| person | 9413* |

(15') [The t-closeness Principle] An equivalence class is said to have t-closeness if the distance between the distribution of a sensitive attribute in this class and distribution of the attribute in the whole table is no more than a threshold t. A table is said to have t-closeness if all equivalence classes have t-closeness.

(a) Given the anonymized table (table 3), where the quasi-identifier attributes are *ZIP Code* and *Age* and the sensitive attribute is *Salary*. Please give the value of $t$ so that table 3 satisfies $t$-closeness. Please use **Earth Mover's distance (EMD)** to calculate the distance between two distributions.

**Hint.** The overall distribution of the Income attribute is $\boldsymbol{Q} = \{3k, 4k, 5k, 6k, 7k, 8k, 9k, 10k, 11k\}$ (We use the notation $\{v_1, v_2, \cdots, v_m\}$ to denote the uniform distribution where each value in $\{v_1, v_2, \cdots, v_m\}$ is equally likely.) The first equivalence class in table 3 has distribution $\boldsymbol{P}_1 = \{3k, 5k, 9k\}$.

[Earth Mover's distance (EMD)]. The *Salary* is the numerical attribute. Numerical attribute values are ordered. Let the attribute domain be $\{v_1, v_2, \cdots, v_m\}$, where $v_i$ is the $i^{th}$ smallest value. Let $\boldsymbol{P} = \{p_1, p_2, \cdots, p_m\}$ and $\boldsymbol{Q} = \{q_1, q_2, \cdots, q_m\}$ be distributions. we use *Ordered Distance* to calculate the distance between two values. Let $r_i = p_i - q_i (i = 1, 2, \cdots, m)$, then EMD between $\boldsymbol{P}$ and $\boldsymbol{Q}$ can be calculate as:

$$
\begin{aligned}
D[\boldsymbol{P}, \boldsymbol{Q}] &= \frac{1}{m-1}(|r_1| + |r_1 + r_2| + \cdots + |r_1 + r_2 + \cdots + r_{m-1}|) \\
&= \frac{1}{m-1} \sum_{i=1}^{m} |\sum_{j=1}^{i} r_j|
\end{aligned} \tag{1}
$$

[Ordered Distance] *Ordered Distance* between two values is based on the number of values between them in the total order, i.e., $ordered\_list(v_i, v_j) = \frac{|i-j|}{m-1}$.

| ZIP Code | Age | Salary |
|---|---|---|
| 4767* | $\leq 40$ | 3K |
| 4767* | $\leq 40$ | 5K |
| 4767* | $\leq 40$ | 9K |
| 4790* | $\geq 40$ | 6K |
| 4790* | $\geq 40$ | 11K |
| 4790* | $\geq 40$ | 8K |
| 4760* | $\leq 40$ | 4K |
| 4760* | $\leq 40$ | 7K |
| 4760* | $\leq 40$ | 10K |

Table 3: The anonymized table.

$m = 9$
$sensitive\ \ attribute\ \ domain : Q = \{3k, 4k, 5k, 6k, 7k, 8k, 9k, 10k, 11k\}$, $p_q = (1/9, 1/9, 1/9, 1/9, 1/9, 1/9, 1/9, 1/9, 1/9)$
$P_1 = \{3k, 5k, 9k\}, rank = \{1, 3, 7\}, p_{1w} = (1/3, 1/3, 1/3)$
$P_2 = \{6k, 8k, 11k\}, rank = \{4, 6, 9\}, p_{2w} = (1/3, 1/3, 1/3)$
$P_3 = \{4k, 7k, 10k\}, rank = \{2, 5, 8\}, p_{3w} = (1/3, 1/3, 1/3)$
对于$3k$，将其分别以$\frac{1}{9}$的概率搬到相近的$3k, 4k, 5k$。此处参考：
《$t - Closeness : PrivacyBeyondk - Anonymityandl - Diversity$》页脚注释4

$$
\begin{aligned}
D[P_1, Q] &= \frac{1}{m-1}(|r_1| + |r_1 + r_2| + \ldots + |r_1 + r_2 + \ldots + r_{m-1}|) \\
&= \frac{1}{8} * \frac{1}{9}(|1-1| + |1-2| + |1-3| + |3-4| + |3-5| + |3-6| + |7-7| + |7-8| + |7-9|) \\
&= \frac{1}{72}(0+1+2+1+2+3+0+1+2) \\
&= \frac{1}{6}
\end{aligned}
$$

$$
\begin{aligned}
D[P_2, Q] &= \frac{1}{m-1}(|r_1| + |r_1 + r_2| + \ldots + |r_1 + r_2 + \ldots + r_{m-1}|) \\
&= \frac{1}{8} * \frac{1}{9}(|4-1| + |4-2| + |4-3| + |6-4| + |6-5| + |6-6| + |9-7| + |9-8| + |9-9|) \\
&= \frac{1}{72}(3+2+1+2+1+0+2+1+0) \\
&= \frac{1}{6}
\end{aligned}
$$

$$
\begin{aligned}
D[P_3, Q] &= \frac{1}{m-1}(|r_1| + |r_1 + r_2| + \ldots + |r_1 + r_2 + \ldots + r_{m-1}|) \\
&= \frac{1}{8} * \frac{1}{9}(|2-1| + |2-2| + |2-3| + |5-4| + |5-5| + |5-6| + |8-7| + |8-8| + |8-9|) \\
&= \frac{1}{72}(1+0+1+1+0+1+1+0+1) \\
&= \frac{1}{12}
\end{aligned}
$$

$t = max(D[P_1, Q], D[P_2, Q], D[P_3, Q]) = 0.167$

**4. (25′) Given the following private table (table 4): Please answer the following questions:**

| Name | Age | Gender | Nationality | Salary | Condition |
|---|---|---|---|---|---|
| Ann | 35 | F | Japanese | 40K | Viral Infection |
| Bluce | 27 | M | American | 38K | Flu |
| Cary | 41 | F | India | 45K | Heart Disease |
| Dick | 32 | M | Korean | 38K | Flu |
| Eshwar | 52 | M | Japanese | 61K | Heart Disease |
| Fox | 22 | M | American | 22K | Flu |
| Gary | 36 | M | India | 34K | Flu |
| Helen | 26 | F | Chinese | 26K | Cancer |
| Irene | 18 | F | American | 16K | Viral Infection |
| Jean | 25 | F | Korean | 38K | Cancer |
| Ken | 38 | M | American | 55K | Viral Infection |
| Lewis | 47 | M | American | 64K | Heart Disease |
| Martin | 24 | M | American | 37K | Viral Infection |

Table 4: Private table.

**(a) (5′) Given the health condition as the sensitive attribute, please name the quasi-identifier attributes.**

quasi-identifier attributes: Age,Gender,Nationality,Salary

**(b) (15′) Let the valid range of age be {0, · · · , 120}. Given the health condition as the sensitive attribute, design a cell-level generalization solution to achieve k-Anonymity, where k = 2. Please give the generalization hierarchies, released table and calculation of the loss metric (LM) of your solution.**

$$A_1 = \{[18-29],[30-55]\}$$

$$\uparrow$$

$$age\ domain : A_0 = \{18, 22, 24, 25, 26, 27, 32, 35, 36, 38, 41, 47, 52\}$$

$$S_1 = \{[16K-39K],[40K-65K]\}$$

$$\uparrow$$

$$S_0 = \{16K, 22K, 26K, 34K, 37K, 38K, 40K, 45K, 55K, 61K, 64K\}$$

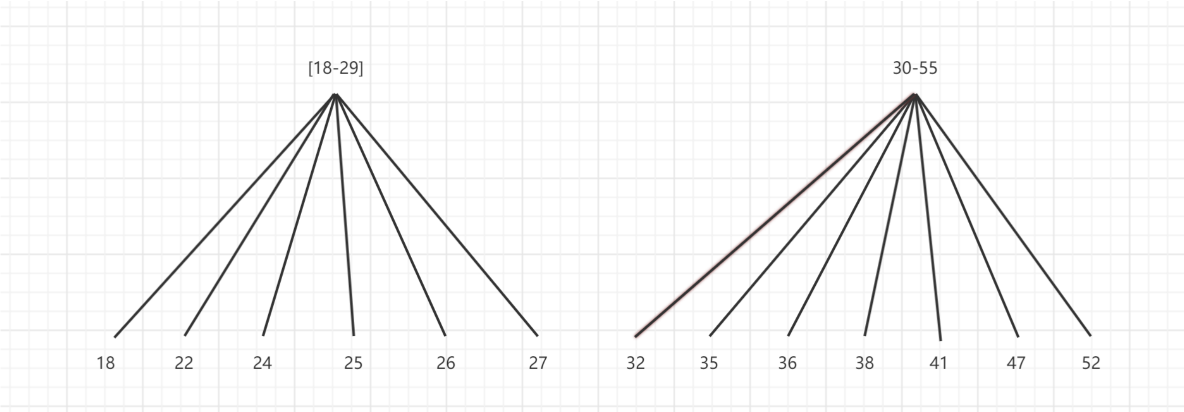
由表中$Gender = F, Nationality = American$只有一个元组

要保证2匿名，所有$Nationality$属性都需泛化成$Person$

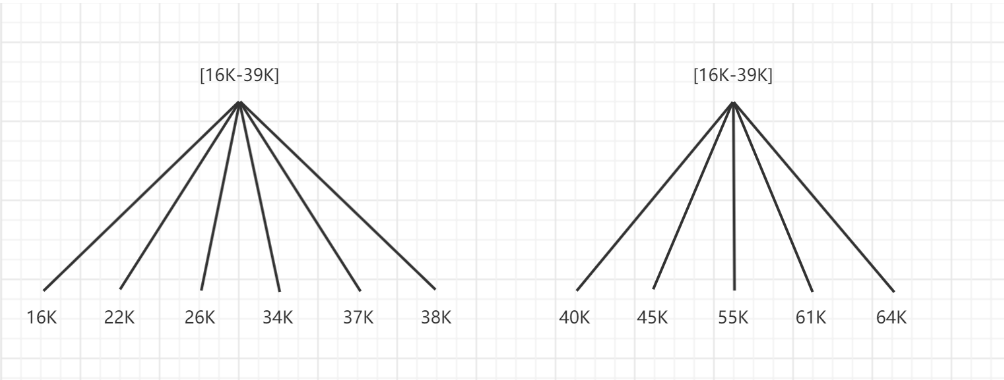
$$N_1 = \{Person\}$$

$$\uparrow$$

$$N_0 = \{American, Chinese, Japanese, Korean, Indian\}$$




Age generalization hierarchies




Salary generalization hierarchies

| Age | Gender | Nationality | Salary(K) | Condition |
|---|---|---|---|---|
| [30-55] | F | Person | [40-65] | Viral Infection |
| [18-29] | M | Person | [16-39] | Flu |
| [30-55] | F | Person | [40-65] | Heart Disease |
| [30-55] | M | Person | [16-39] | Flu |
| [30-55] | M | Person | [40-65] | Heart Disease |
| [18-29] | M | Person | [16-39] | Flu |
| [30-55] | M | Person | [16-39] | Flu |
| [18-29] | F | Person | [16-39] | Cancer |
| [18-29] | F | Person | [16-39] | Viral Infection |
| [18-29] | F | Person | [16-39] | Cancer |
| [30-55] | M | Person | [40-65] | Viral Infection |
| [30-55] | M | Person | [40-69] | Heart Disease |
| [18-29] | M | Person | [40-65] | Viral Infection |

**染色图如下，满足2-Anonymity要求:**

| Age | Gender | Nationality | Salary(K) | Condition |
|---|---|---|---|---|
| [30-55] | F | Person | [40-65] | Viral Infection |
| [18-29] | M | Person | [16-39] | Flu |
| [30-55] | F | Person | [40-65] | Heart Disease |
| [30-55] | M | Person | [16-39] | Flu |
| [30-55] | M | Person | [40-65] | Heart Disease |
| [18-29] | M | Person | [16-39] | Flu |
| [30-55] | M | Person | [16-39] | Flu |
| [18-29] | F | Person | [16-39] | Cancer |
| [18-29] | F | Person | [16-39] | Viral Infection |
| [18-29] | F | Person | [16-39] | Cancer |
| [30-55] | M | Person | [40-65] | Viral Infection |
| [30-55] | M | Person | [40-65] | Heart Disease |
| [18-29] | M | Person | [16-39] | Viral Infection |

$$
\begin{aligned}
&\text{对于}\,Age(\text{数值}):\\
&\quad T[18-29]=\frac{29-18}{55-18}=\frac{9}{37}\\
&\quad T[30-55]=\frac{55-30}{55-18}=\frac{25}{37}\\
&\quad LM_{age}=(6*\frac{9}{37}+7*\frac{25}{37})/13=\frac{229}{481}\\
&\text{对于}\,Salary(\text{数值}):\\
&\quad T[16-39]=\frac{39-16}{65-16}=\frac{23}{49}\\
&\quad T[40-65]=\frac{65-40}{65-16}=\frac{25}{49}\\
&\quad LM_{Salary}=(8*\frac{23}{49}+5*\frac{25}{49})/13=\frac{309}{637}\\
&\text{对于}\,Nationality(\text{记录}):\\
&\quad T[American]=\frac{|M|-1}{|A|-1}=\frac{5-1}{5-1}=1\\
&\quad T[Chinese]=\frac{|M|-1}{|A|-1}=\frac{5-1}{5-1}=1\\
&\quad T[Japanese]=\frac{|M|-1}{|A|-1}=\frac{5-1}{5-1}=1\\
&\quad T[Korean]=\frac{|M|-1}{|A|-1}=\frac{5-1}{5-1}=1\\
&\quad T[Indian]=\frac{|M|-1}{|A|-1}=\frac{5-1}{5-1}=1\\
&\quad LM_{Nationality}=1\\
&LM=LM_{age}+LM_{Salary}+LM_{Nationality}=1.96
\end{aligned}
$$

## (c) (5’) Please design a k-anonymization algorithm to optimize the loss metric.

```
先对所有准标识符分别构建generalization hierarchies
将所有准标识符的泛化方式作为向量分量，构建lattice，记最高高度为h
DFS(lattice)中每一种泛化方式，是否满足K匿名，并记录LM值
在所有LM值中选择最小的，其对应的泛化方式即是最优的K匿名方式
```

5. (20’) Suppose that private information $x$ is a number between 0 and 1000. This number is chosen as a random variable $X$ such that 0 is 1%-likely whereas any non-zero is only about 0.1%-likely:

$$P[X=0]=0.01, P[X=k]=0.00099, k=1\cdots 1000 \qquad (2)$$

Suppose we want to randomize such a number by replacing it with a new random number $y=R(x)$ that retains some information about the original number $x$. Here are three possible methods to do it:

(a) Given $x$, let $R_1(x)$ be $x$ with 20% probability, and some other number (chosen uniformly at random in $\{0,\cdots,1000\}$) with 80% probability.

(b) Given $x$,let $R_2(x)$ be $(x+\delta) mod\ 1001$, where $\delta$ is chosen uniformly at random in $\{-100\cdots 100\}$.

(c) Given $x$, let $R_3(x)$ be $R_2(x)$ with 50% probability, and a uniformly random number in $\{0,\cdots,1000\}$ otherwise.

Please answer the following questions:

(a) (15’) Compute prior and posterior probabilities of two properties of $x$: 1) $X=0$; 2) $x\in\{200,\cdots,800\}$ using the above three methods respectively.

(b) (5’) Which method is better? Why?

### (a)

$$\text{先验概率为} P_{i-\text{先}} \quad i=1,2,3$$

$$\text{后验概率为} P_{i-\text{后}} \quad i=1,2,3$$

$$X=0:$$

$$\begin{aligned}
P_{1-\text{先}}(X=0) &= P(X=0)=0.01 \\
P(R_1(X)=0) &= 0.2*P(X=0)+0.8*P(Z=0|Z\in[0,1000])=0.0027992 \\
P_{1-\text{后}} &= P(X=0|R_1(X)=0) \\
&= \frac{P(X=0,R_1(X)=0)}{P(R_1(X)=0)} \\
&= \frac{P(X=0)*P(R_1(X)=0|X=0)}{P(R_1(X)=0)} \\
&= \frac{0.01*(0.2+0.8*\frac{1}{1001})}{0.0027992}=0.717
\end{aligned}$$

$$P_{2-先}(X=0)=P(X=0)=0.01$$

$$P(R_2(X)=0)|X=0)=P(\delta=0)=\frac{1}{201}$$

$$\begin{aligned}P(R_2(X)=0)=P\{(X+\delta)\equiv 0(mod\ \ 1001)|\delta\in U[-100,100]\}\\ &=P(X+\delta=0)+P(X+\delta=1001)\\ &=P\{X=0,\delta=0\}+P\{X=i,\delta=-i\}(i\in[1,100])+P\{X=1001-i,\delta=i\}(i\in[1,100])\\ &=0.01*\frac{1}{201}+100*0.00099*\frac{1}{201}+100*0.00099*\frac{1}{201}\\ &=\frac{0.208}{201}=0.0010348258706\end{aligned}$$

$$\begin{aligned}P_{2-后}&=P(X=0|R_2(X)=0)\\ &=\frac{P(X=0,R_2(X)=0)}{P(R_2(X)=0)}\\ &=\frac{P(X=0)*P(R_2(X)=0|X=0)}{P(R_2(X)=0)}\\ &=\frac{0.01*\frac{1}{201}}{0.0010348}=0.048078\end{aligned}$$

$$P_{3-先}(X=0)=0.01$$

$$\begin{aligned}P(R_3(x)=0)=0.5*P(R_2(X)=0)+0.5*P(Z=0|Z\in U[0,1000])\\ &=\frac{0.104}{201}+0.5*\frac{1}{1001}=0.0010169134348\end{aligned}$$

$$P(R_3(X)=0)|X=0)=0.5*P(R_2(X)=0)|X=0)+0.5*\frac{1}{1001}=0.002987$$

$$\begin{aligned}P_{3-后}&=P(X=0|R_3(X)=0)\\ &=\frac{P(X=0,R_3(X)=0)}{P(R_3(X)=0)}\\ &=\frac{P(X=0)*P(R_3(X)=0|X=0)}{P(R_3(X)=0)}\\ &=0.02937\end{aligned}$$

$X \in [200, 800]:$

$$P_{1-先}(X \in [200, 800]) = P(X \in [200, 800]) = 601 * 0.00099 = 0.59499$$

$$P(R_1(X) = 0 | X \in [200, 800])$$

$$= 0.8 * \frac{1}{1001} = \frac{0.8}{1001}$$

$$P(R_1(X) = 0)$$

$$= 0.2 * 0.01 + 0.8 * \frac{1}{1001} = 0.0027992$$

$$P_{1-后} = P(X \in [200, 800] | R_1(X) = 0)$$

$$= \frac{P(X \in [200, 800], R_1(X) = 0)}{P(R_1(X) = 0)}$$

$$= \frac{P(X \in [200, 800]) * P(R_1(X) = 0 | X \in [200, 800])}{P(R_1(X) = 0)}$$

$$= 0.169875$$

$$P_{2-先}(X \in [200, 800]) = 0.59499$$

$$P(R_2(X) = 0 | X \in [200, 800]) = 0$$

$$P(R_2(X) = 0) = P\{(X + \delta) \equiv 0 (mod \ \ 1001) | \delta \in U[-100, 100]\}$$

$$= P(X + \delta = 0) + P(X + \delta = 1001)$$

$$= P\{X = 0, \delta = 0\} + P\{X = i, \delta = -i\}(i \in [1, 100]) + P\{X = 1001 - i, \delta = i\}(i \in [1, 100])$$

$$= 0.01 * \frac{1}{201} + 100 * 0.00099 * \frac{1}{201} + 100 * 0.00099 * \frac{1}{201}$$

$$= \frac{0.208}{201} = 0.0010348258706$$

$$P_{2-后} = P(X \in [200, 800] | R_2(X) = 0$$

$$= \frac{P(X \in [200, 800], R_2(X) = 0)}{P(R_2(X) = 0)}$$

$$= \frac{P(X \in [200, 800]) * P(R_2(X) = 0 | X \in [200, 800])}{P(R_2(X) = 0)}$$

$$= 0$$

$$P_{3-先}(X \in [200, 800]) = 0.59499$$

$$P(R_3(X) = 0) = 0.5 * P(R_2(X) = 0) + 0.5 * \frac{1}{1001} = 0.0010169$$

$$P(R_3(X) = 0 | X \in [200, 800])$$

$$= 0.5 * P(R_2(X) = 0 | X \in [200, 800]) + 0.5 * \frac{1}{1001} = 0.0004995$$

$$P_{3-后} = P(X \in [200, 800] | R_3(X) = 0)$$

$$= \frac{P(X \in [200, 800], R_3(X) = 0)}{P(R_3(X) = 0)}$$

$$= \frac{P(X \in [200, 800]) * P(R_3(X) = 0 | X \in [200, 800])}{P(R_3(X) = 0)}$$

$$= 0.292258$$

| Given | X=0 | X∈[200,800] |
| --- | --- | --- |
| R1(X)=0 | 71.7% | 16.9875% |
| R2(X)=0 | 4.8% | 0 |
| R3(X)=0 | 2.9% | 29.2258% |
| nothing | 1% | 59.499% |

**(b): 选择R3好，根据前述计算（结果如上表），X是需要保护的数据，对于R3而言，每种情况的后验概率都是与先验概率最接近的（之差最小）。也就是说，对X采用R3的方法处理之后，隐私泄露最少。而其他的方法，如R2，在已知R2=0的情况下，X∈[200,800]的概率为0，得到了确定性信息，这是严重的隐私泄露。故而R3最好。**

6. (15') [$(\alpha, \beta)$-Privacy] Let $R$ be an algorithm that takes as input $u \in D_U$ and outputs $v \in D_V$. $R$ is said to allow an upward $(\alpha, \beta)$-privacy breach with respect to a predicate $\phi$ if for some probability distribution $f$,

$$\exists u \in D_U, \exists v \in D_V \ s.t. \ P_f(\Phi(u)) \le \alpha \ and \ P_f(\Phi(u)|R(u) = v) \ge \beta \quad (3)$$

Similarly, $R$ is said to allow a downward $(\alpha, \beta)$-privacy breach with respect to a predicate $\Phi$ if for some probability distribution $f$,

$$\exists u \in D_U, \exists v \in D_V \ s.t. \ P_f(\Phi(u)) \ge \alpha \ and \ P_f(\Phi(u)|R(u) = v) \le \beta \quad (4)$$

$R$ is said to satisfy $(\alpha, \beta)$-privacy if it does not allow any $(\alpha, \beta)$-privacy breach for any predicate $\Phi$. The necessary and sufficient conditions for $R$ to satisfy $(\alpha, \beta)$-privacy for any prior distribution and any property $\phi$: $\gamma$-amplifying

$$\forall v \in D_V, \forall u_1, u_2 \in D_U, \frac{P(R(u_1) = v)}{P(R(u_2) = v)} \le \gamma \quad (5)$$

(a) Let $R$ be an algorithm that is $\gamma$-amplifying. Please proof that $R$ does not permit an $(\alpha, \beta)$-privacy breach for any adversarial prior distribution if

$$\gamma \le \frac{\beta}{\alpha} \frac{1-\alpha}{1-\beta}. \quad (6)$$

说明，(4)式$\alpha, \beta$互换，(6)应为严格小于!!!

$pf$：可知$\forall u \in D_U, p(R(u) = v) > 0$，若否，则$\gamma \to \infty$，令随机变量$V = R(U)$

考虑$u$的任意分布$p_U$，至少在一个$u \in D_U$上，$p_U > 0$，故：

$$P[V = v] = P[U = v] * p(R(u) = v) > 0$$

反证法：假设对$\phi(u)$，存在$(\alpha, \beta) - privacy\ breach$，因为由(3)式，则$\phi(u)$不可能对所有$u \in D_U$均为真。

$P[\phi(U)] \le \alpha < 1$。同样，由(3)，$P[\phi(U)|Y = y] \ge \beta > 0$。

故而令$u_1 \in \{u \in D_u | \phi(u) = 1, p[R(u) = v] = max_{\phi(u')}\ p[R(u') = v]\}$

$u_2 \in \{u \in D_u | \phi(u) = 0, p[R(u) = v] = min_{\neg\phi(u')}\ p[R(u') = v]\}$

即$u_1$是使$\phi(u)$为真切最大概率映射到$v$的值，$u_2$是使$\phi(u)$为假且最小概率映射到$v$的值。

$$P[\phi(U)|V = v] = \sum_{\phi(u)} P[U = u|V = v] = \sum_{\phi(x)} \frac{P[U = u] * p[R(u) = v]}{P[V = v]}$$

$$\le \frac{p[R(u_1) = v]}{P[V = v]} * \sum_{\phi(u)} P[U = u] = p[R(u_1) = v] * \frac{P[\phi(U)]}{P[V = v]}$$

同样

$$P[\neg\phi(U)|V = v] = \sum_{\neg\phi(u)} P[U = u|V = v] = \sum_{\neg\phi(x)} \frac{P[U = u] * p[R(u) = v]}{P[V = v]}$$

$$\ge \frac{p[R(u_2) = v]}{P[V = v]} * \sum_{\neg\phi(u)} P[U = u] = p[R(u_2) = v] * \frac{P[\neg\phi(U)]}{P[V = v]}$$

我们知道$P[\phi(U)|V = v] \ge \beta > 0, P[\phi(U)] > 0$，所以：

$$\frac{P[\neg\phi(U)|V = v]}{P[\phi(U)|V = v]} \ge \frac{p[R(u_1) = v]}{p[R(u_2) = v]} * \frac{P[\neg\phi(U)]}{P[\phi(U)]}$$

因为$R(u)$至多是$v$的$\gamma$倍，所以

$$\frac{1 - P[\phi(U)|V = v]}{P[\phi(U)|V = v]} \ge \frac{1}{\gamma} * \frac{1 - P[\phi(U)]}{P[\phi(U)]}$$

又有：

$$\frac{1 - P[\phi(U)|V = v]}{P[\phi(U)|V = v]} \le \frac{1-\beta}{\beta}; \frac{1 - P[\phi(U)]}{P[\phi(U)]} \ge \frac{1-\alpha}{\alpha}$$

由上可得，$\frac{1-\beta}{\beta} \ge \frac{1}{\gamma} * \frac{1-\alpha}{\alpha}$，即$\frac{\beta}{\alpha} * \frac{1-\alpha}{1-\beta} \le \gamma$，与(6)矛盾，故不能满足$upward\ (\alpha, \beta) - privacy\ breach$。

对于$downward\ (\alpha, \beta) - privacy\ breach$，令$\alpha' = 1 - \beta, \beta' = 1 - \alpha$，转化为$upward\ (\alpha', \beta') - privacy\ breach$的证明，同理可证。

或将$\alpha' = 1 - \beta, \beta' = 1 - \alpha$带入$upward$已得证的不等式中：

$$\frac{\beta'}{\alpha'} * \frac{1-\alpha'}{1-\beta'} = \frac{1-\alpha}{1-\beta} * \frac{\beta}{\alpha} > \gamma$$

证毕。