基于隐私保护的ID匹配

1. 隐私集合求交（PSI）

> 关键词：
>
> PSI （Private Set Intersection）：隐私集合求交
>
> MPC（Secure Multi-Party Computation）: 多方安全计算

2. RSA原理

2.1. 数学基础

> 定理一： 如果两个正整数a和n互质，那么一定可以找到整数b，使得 ab-1 被n整除，或者说ab被n除的余数是1。
>
> 这时，b就叫做a对模数n的**"模反元素"**。即：
>
> $ab \quad \% \quad n = 1$
>
> 定理二： 已知n为两质数的乘积，$\phi(n)$为n的欧拉函数，e为小于$\phi(n)$且与$\phi(n)$互质的整数，d为e对模数$\phi(n)$的翻模元素，则对于任意小于n且大于1的正整数m：
>
> $m^{ed} \quad \% \quad n = m$
>
> 定理三：唯一分解定理：任一大于1的自然数，要么本身是质数，要么可以分解为几个质数之积，且这种分解是唯一的。

2.2. 密钥生成

- 选不相等的两质数`p，q`
- $n = p * q$
- 计算n的欧拉函数 $\phi(n) = (p-1)(q-1)$
- 选整数`e:` $1 < e < \phi(n)$ ，且e与$\phi(n)$互质
- 计算e对于模数$\phi(n)$的模反元素d：ed 整除 $\phi(n)$余数是1 ，根据定理一，d一定存在

$$ed = 1(\mod \phi(n))$$

最后有p,q,n,$\phi(n)$,e,d，只有n,e是公开的

公钥(n,e) 只能加密小于n的整数m，那么如果要加密大于n的整数，该怎么办？有两种解决方法：一种是把长信息分割成若干段短消息，每段分别加密；另一种是先选择一种"对称性加密算法"（比如 **DES** ），用这种算法的密钥加密信息，再用RSA公钥加密DES密钥。

2.3. 加密

被加密的整数m必须是小于n的整数，加密成整数c

$$m^e \ \% \ n = c$$

2.4. 解密

整数c解密为整数m

$$c^d \ \% \ n = m$$

2.5. RSA可靠性

（1）ed≡1 (mod φ(n))。只有知道e和φ(n)，才能算出d。

（2）φ(n)=(p-1)(q-1)。只有知道p和q，才能算出φ(n)。

（3）n=pq。只有将n因数分解，才能算出p和q。

重点在于$\phi(n)$不能被泄露

3. 基于RSA和HASH的ID匹配

3.1. 前期准备

企业B生成密钥：

1. 选**互质**的两个正整数p,q, 计算n=p*q；n的欧拉函数$\phi(n)$=(p-1)*(q-1)

注：

n要大于数据量，n转换为二进制的位数就是RSA加密位数

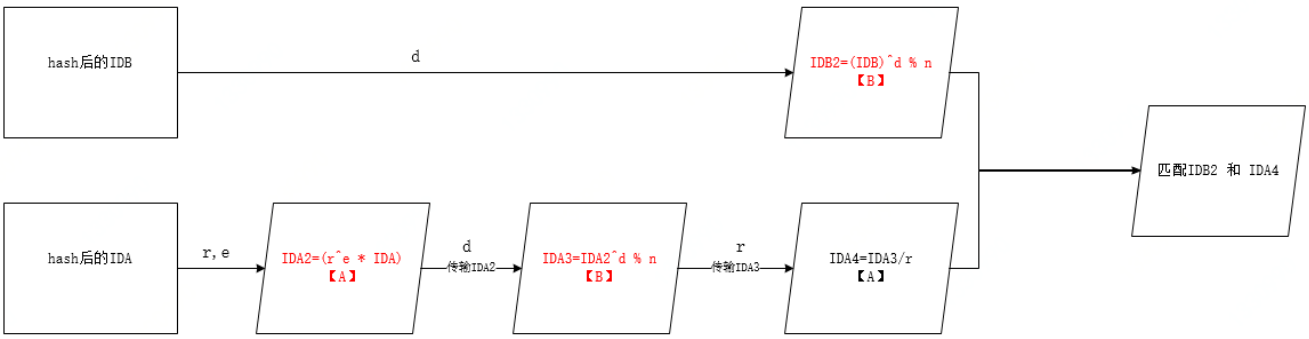
假设数据量条数换算成二进制是m位，通常选择p,q 使得 $2^m < n < 2^{m+1}$ （1亿用户m=27，可选择28位加密）

2. 选择满足条件的正整数e: $1 < e < \phi(n)$ ，且e与$\phi(n)$互质
3. 计算e对于模数$\phi(n)$的模反元素d

企业A为每个用户生成随机整数r, **r为小于n的整数**，可用数据打乱后的行号代替

注：企业A也可选择几个指定的r为其全量用户加密

3.2. 数据加密过程



注释：

- 【】表示操作方
- ID_b : 一次幂乘，一次求余数
- ID_a : 两次幂乘，一次求余，一次除法