# 题目介绍

一共有2个题

任务1：计算黑名单交集

半诚实双方，网络带宽100Mbps

双方都是100万数量级

求交字段是身份证，18位int

任务2：根据ID匿踪查询黑名单信息

半诚实双方，网络带宽100Mbps

待查寻10条（18位），被查询100万（身份证、日期、年龄）

# 方案设计和实现

项目要求和对应解决思路：

安全通信：secretflow通信采用brpc，自带mTLS认证

针对比赛环境的优化：选择通信量、计算量都是中等的方案

算法的安全性（安全参数、随机熵源安全、密钥安全）：密钥都是临时的

PSI安全证明是LPN assumption, 安全参数选择128位、40位

PIR安全证明是OMGDH assumption in ROM，安全参数选择128位

Baseline是sf.spu中的相关函数调用时长

## PSI方案

目前有以下四种技术路线：

1. 基于公钥体系（计算开销过大）

DH-PSI, RSA-PSI, NR-PSI

1. 基于电路（主要用于计算交集的统计函数）

[HDK12],[PSW18],[PP21]

1. 基于同态加密（主要用于非平衡场景）

[FNPO4],[CLR17],[CMG21]

1. 基于不经意传输协议（目前两方平衡场景下最高效的方案）

[KKRT16],[RR22],[BC23]

一些OTE PSI的信息：（这是已经发展了十多年的技术路线）

|  |  |  |  |
| --- | --- | --- | --- |
| 协议 | 敌手模型 | 核心技术/数据结构 | OTE类型 |
| DCW13 | 半诚实、恶意 | （混淆）布隆过滤器 | IKNP03 |
| KKRT16 | 半诚实 | Cuckoo hash，伪随机码 | KK13 |
| RR22 | 恶意 | Subfield VOLE, OKVS | CRR21 |
| BC23 | 恶意 | Subfield VOLE, hash-based | CRR21 |

视频中选择了BC23，hash-based是Cuckoo hash的变种，主要步骤如下

1. 接收者用(d,k)布谷鸟哈希将自己的输入元素映射到一张表中。其中,k 表示哈希函数的个数，d 表示该表每一行所能容纳的最多的元素个数。
2. 接收者和发送者运行一个 BaRKOPRF协议。BaRK-OPRF基于VOLE构造。
3. 发送者计算k·n个PRF:

Ht={PRF(ht(y)|y∈Y},其中t∈[1,k];发送者将Ht发送给接收者

1. 对于接收者的输入元素x∈X，假设x 是被哈希函数 h 映射到了第i个桶的第j个位置，则接收者需要检査x对应的PRF 值是否出现在了集合Ht中。如果出现了，则输出x。

## PIR方案

方案对比

索引稀疏型：检索ID跳跃比较大

索引连续型：例如1、2、3、4、5

题目要求保护服务器隐私

* 技术选型：Asymmetric Labeled PSI, 解释：

1. Labeled: 服务器和客户端在ID 列上做交集运算;服务器将交集ID 对应的 Label列返回给客户端
2. Asymmetric: 服务器拥有100万条数据，而客户端只有10条数据

构造Labeled PSI：

PSI内部都有OPRF，选两把key，分别用来对比id、加密label

对于非平衡数据集特殊处理：

[HKZ18]为了非平衡数据集下的亚线性通信效率，引入了基于同态加密的不经意多项式求值

（同态、多项式插值都吃计算资源）

测试环境的带宽达到了 100Mbps，没有必要用计算换通信

所以视频构造了Labeled ECDH PSI, 对label做了一些扩展

方案调用：secretflow/spu: ecdh\_oprf\_psi

通过PSI接口直接调用

结论：针对环境定制方案，是提升性能的关键

# 总结

1. 高安全

方案设计:具有形式化安全性证明

方案实现:mTLS,保证通信安全;InteIDRNG熵源保证随机熵源安全;临时密钥保证密钥安全

1. 性能调优

定制的优化方案:为比赛测试环境设计专有方案，将任务二 baseline 的运行效率提升了 10 倍

工程优化:(1) SIMD 指令优化;(2)FourQ椭圆曲线加速

1. 工程化实现程度高

相关方案实现已集成到 secretflow/spu模块中,遵循 secretflow的协议接口规范，可使用 python脚本轻松部署