PSU

No Author Given

No Institute Given

摘要 PSU

Keywords: PSU

1 Introduction

本文主要对当前 PSU 协议的通信复杂度进行优化.

我们试图考虑基于全同态加密（FHE）将 PSU 的通信复杂度降到 $O(|Y|)$，其中 Y 表示 R 的集合大小，在 unbalance 场景下，当 Y 很小时，通信量较小。

思路:

1. 考虑之前和 Frikken-07 想法重合的技术，基于加法同态加密（HE）设计 PSU，通信复杂度为 $O(|X|)+O(|Y|)$：因为 S 需要把自己的集合 $X=\{x_1,x_2,\cdots,x_n\},|X|=n$ 表示集合元素个数；计算多项式 $y=f(x),s.t.f(x_i)=0$. 然后将多项式系数分别用自己的公钥加密发送给 R; 此时 R 可以计算自己的元素 y_i 对应的多项式的密文 $E(f(y_i))$, 并通过随机数混淆，计算 $c_1=E(rf(y_i))$ 和 $c_2=E(rf(y_i)y_i)$ 并发送给 S，[此处需要对 c_2 进行重加密，不然 S 可以暴力破解 y_i，如果 Y 的空间较小]；然后，S 可解密获得 $m_1=rf(y_i),m_2=rf(y_i)y_i$，如果 $m_1=m_2=0$,S 知道 $y_i\in X$，但不知道 y_i，如果 $m_1\neq 0,m_2\neq 0$，则 S 知道 $y_i\notin X$, 且可计算 $y_i=m_2/m_1$.

分析 Frikken-07 协议：该协议只需要 |X|+2|Y| 个通信量，不需要使用 OT 协议来选择集合中的非交集部分的元素，其中 |X| 表示发送方 S 的集合大小，|Y| 表示接收方 R 的集合大小。该协议主要的通信量在第一轮，S 要把多项式各个系数加密发给 R（目的是要 R 能同态计算出 $E(f(y_i))$），该部分和 S 的集合的大小是一致的。

2. 我们考虑类似对偶的想法，让多项式的密文 $E(f(y_i))$ 由 S 来计算? 初步考虑加法同态加密的方式，即 R 发送 $E(y_i),E(y_i^2),\cdots,E(y_i^{n-1})$ 给 S，S 可计算 $E(f(y_i))$!!![这样思考的动机在于：这步由 S 计算则不需要发送多项式，可以考虑 FHE 优化这一步的通信量]

这个想法存在以下问题:

1) 加密的公钥如果是 S 的，则泄漏了 R 的信息 y_i；如果公钥是 R 的，则后续可能会泄漏 S 的信息，比如 n 次询问就泄漏了多项式，而且如果其中有 0，R 就知道了交集;
2) 如果 R 的集合 Y 很大，R 第一轮发送的消息的通信量很大，具体为：$|X||Y|$.

对上述对偶想法的改进过程：首先如果公钥是 S 的肯定行不通，因为完全泄漏了 R 的信息!!!

1) 初步改进：公钥是 R 的，S 选择随机数 r 用于隐藏 $f(y_i)$ 的信息，返回 $E(r+f(y_i))$，然后 R 解密出 $r+f(y_i)$ 返回给 S；S 判断如果 $r+f(y_i)=r$，则说明 $y_i \in X$，否则 S 知道 $y_i \notin X$，且获得 $f(y_i)$；
   如此设计导致多加一次交互，还需要使用一次 OT 协议，同时如果 R 的集合很大的时候，通信量反而增加了，变成 $n|Y|+|Y|+|Y|$;
2) 进一步改进：上述 OT 的使用，是因为 S 难求解多项式的根。因为解密方是 R，所以 S 需要选择随机数保护 $f(y_i)$，防止泄漏多项式信息；针对这个问题，我们可以考虑 Two-party HE，防止 R 解密泄漏信息；如此，R 输出解密分享，然后 S 可完整解密，剩下部分和 Frikken-07 类似；
   这个改进可以不需要 OT，但是第一轮的通信量依旧较大！
3) 再进一步改进：因为使用 HE，S 需要知道 y_i 的所有次方的密文才能计算多项式的密文 $E(f(y_i))$；所以 R 需要公布 $E(y_i),E(y_i^2),\cdots,E(y_i^{n-1})$；即第一轮的通信量是 $|X||Y|$，通讯量依旧较大。
   针对这个问题：可考虑两方全同态加密 TP-FHE：R 发送 $FHE(y_i)$ 即可，S 可直接计算 $TP-FHE(f(y_i))$，R 可选择随机数 r_i, 计算 $c_{1i}=TP-FHE(r_if(y_i)), c_{2i}=TP-FHE(r_if(y_i)y_i)$(这里的 r 是防止泄漏 y_i 的信息 (穷搜))；并计算解密分享 $d_{R1i}=\mathsf{Dec}_{sk_R}(c_{1i})$ 和 $d_{R2i}=\mathsf{Dec}_{sk_R}(c_{2i})$;R 发送 $c_{1i},c_{R2i},d_{1i},d_{R2i}$ 给 S; S 解密 $d_{S1i}=\mathsf{Dec}_{sk_S}(c_{1i})$ 和 $d_{S2i}=\mathsf{Dec}_{sk_S}(c_{2i})$, 然后运行组合算法获得 $m_{1i}=\mathsf{Combine}(d_{R1i},d_{S1i})=r_if(y_i)$ 和 $m_{2i}=\mathsf{Combine}(d_{R2i},d_{S2i})=r_if(y_i)y_i$, 如果 $m_1=m_2=0$,S 知道 $y_i\in X$，但不知道 y_i，如果 $m_1\neq 0, m_2\neq 0$，则 S 知道 $y_i\notin X$, 且可计算 $y_i=m_2/m_1$.
   但是上述方案计算量太大，可考虑简单优化，R 发送 $FHE(y_i),FHE(y_i^2),\cdots,FHE(y_i^{\log n})$ 即可！

对比总结：

1) Frikken-07 方案的通信轮数 2 轮，通信复杂度 $|X|+2|Y|$
2) TP-FHE 方案的通信轮数 3 轮，通信复杂度 $|Y|+4|Y|$，所有的通信量和 X 无关，如果 Y 很小的话 (unbalance 场景下)，通信效率较高。

2 HE-based PSU (Frikken-07)

$f(x)=a_1+a_2x+\cdots+a_nx^{n-1}$,s.t. for all $i\in[n]$, $f(x_i)=0$

$S(x_i\in X, i\in[n])$		$R(y_i\in Y)$
$f(x)=a_1+a_2x+\cdots+a_nx^{n-1}$	$\xrightarrow{\mathsf{HE}_{pk_S}(a_i),i\in[n]}$	$\mathsf{HE}_{pk_S}(f(y_i))$
$m_{1i}=\mathsf{Dec}_{sk_S}(c_{1i}), m_{2i}=\mathsf{Dec}_{sk_S}(c_{2i})$	$\xleftarrow{c_{1i},c_{2i}}$	$c_{1i}=\mathsf{HE}_{pk_S}(r_if(y_i))$,$c_{2i}=\mathsf{HE}_{pk_S}(r_if(y_i)y_i)$
If $m_{1i}=m_{2i}=0$, $y_i\in X$		
If $m_{1i}\neq 0, m_{2i}\neq 0$, $y_i=m_{2i}/m_{1i}\notin X$		

关键点：R 计算 $c_1=E(rf(y_i))$ 和 $c_2=E(rf(y_i)y_i)$ 并发送给 S，需要对 c_2 进行重加密，不然 S 可以暴力破解 y_i(如果 Y 的空间较小)；

3 HE-based PSU with OT

Frikken-07 对偶的想法，使用 R 的公钥，由 S 计算 $\mathsf{HE}_{pk_R}(f(y_i))$:

具体协议如下：

$S(x_i \in X, i \in [n])$		$R(y_i \in Y)$
$f(x) = a_1 + a_2x + \cdots + a_nx^{n-1}$	$\xleftarrow{\mathsf{HE}_{pk_R}(y_i),\cdots,\mathsf{HE}_{pk_R}(y_i^{n-1})}$	
$r_i \leftarrow \mathbb{Z}_q$,$c_{1i} = \mathsf{HE}_{pk_R}(f(y_i) + r_i)$	$\xrightarrow{c_{1i}}$	$f(y_i) + r_i = \mathsf{Dec}_{sk_R}(c_{1i})$
If $r_i = f(y_i) + r_i$, $y_i \in X$	$\xleftarrow{r_i + f(y_i)}$	
If $r_i \neq f(y_i) + r_i$, $y_i \notin X$		

后续根据 $y_i \notin X$ 还是 $y_i \in X$ 使用 OT 协议选择 y_i;

该协议的通讯量为：$O(|X||Y| + 2|Y|)$+OT 协议的通信量

直觉上当 $r_i \neq f(y_i) + r_i$, $y_i \notin X$ 时，会泄漏 $f(y_i)$ 即 y_i 的信息。但是没有关系，因为在 PSU 中，$y_i \notin X$，S 最后需要获得 y_i 的值。

问题在于，该方法依旧需要 OT，难以使用 Frikken-07 的结构，因为解密的是 R，如果使用乘积的形式 $r_if(y_i)$ 会泄漏交集信息，比如 $r_if(y_i) = 0$，则 R 解密得到 0，即知道 $y_i \in X$.

针对这个问题，我们要保证 R 不能解密！！！所以我们考虑两方解密场景，R 只能提供解密分享，不知道明文，同时第一轮中 S 也无法解密，不会泄漏 R 的信息。

或者考虑在不破坏同态的情况下，修改密文导致 R 不能解密，或者重加密 [目的是 R 不能解密，但能同态运算出 $E(rf(y))$ 和 $E(ryf(y))$]，最后 S 还能解密！！！构造 Frikken-07 的结构 $rf(y)$ 和 $ryf(y)$ 才能不需要 OT，但是如果 R 能解密，如果 $y \in X$，则 R 就知道了，因为解密出来是 0;

4 FHE-based PSU with OT

对 HE-based PSU with OT 使用 FHE 进行扩展，通信量降低为 $O(3|Y|)$+OT 协议的通信量，其中 OT 协议的通信量也是和集合 Y 的大小相关的（S 在 Y 集合中挑元素，与 X 无关），与集合 X 不相关。

具体协议如下：

$S(x_i \in X, i \in [n])$		$R(y_i \in Y)$
$f(x) = a_1 + a_2x + \cdots + a_nx^{n-1}$	$\xleftarrow{\mathsf{FHE}_{pk_R}(y_i)}$	
$r_i \leftarrow \mathbb{Z}_q$,$c_{1i} = \mathsf{FHE}_{pk_R}(f(y_i) + r_i)$	$\xrightarrow{c_{1i}}$	$f(y_i) + r_i = \mathsf{Dec}_{sk_R}(c_{1i})$
If $r_i = f(y_i) + r_i$, $y_i \in X$	$\xleftarrow{r_i + f(y_i)}$	
If $r_i \neq f(y_i) + r_i$, $y_i \notin X$		

5 Two-party HE-based PSU

pk 是 Two-party HE 的公钥，sk_S, sk_R 分别是接收方和发送方的私钥；

$S(x_i \in X, i \in [n])$		$R(y_i \in Y)$
$f(x) = a_1 + a_2x + \cdots + a_nx^{n-1}$	$\xleftarrow{\mathsf{HE}_{pk}(y_i),\cdots,\mathsf{HE}_{pk}(y_i^{n-1})}$	
	$\xrightarrow{c_i=\mathsf{HE}_{pk}(f(y_i))}$	$r_i \leftarrow \mathbb{Z}_q, c_{1i} = \mathsf{HE}_{pk}(r_if(y_i))$
		$c_{2i} = \mathsf{HE}_{pk}(r_iy_if(y_i))$
	$\xleftarrow{c_{1i},c_{2i},d_{R1i},d_{R2i}}$	$d_{R1i} = \mathsf{Dec}_{sk_R}(c_{1i}), d_{R2i} = \mathsf{Dec}_{sk_R}(c_{2i})$
$d_{S1i} = \mathsf{Dec}_{sk_S}(c_{1i}), d_{S2i} = \mathsf{Dec}_{sk_S}(c_{2i})$		
$m_{1i} = \mathsf{Combine}(d_{S1i}, d_{R1i})$		
$m_{2i} = \mathsf{Combine}(d_{S2i}, d_{R2i})$		
If $m_{1i} = m_{2i} = 0$, $y_i \in X$		
If $m_{1i} \neq 0, m_{2i} \neq 0$, $y_i = m_{2i}/m_{1i} \notin X$		

关键点：$c_{1i} = \mathsf{HE}_{pk}(r_if(y_i)), c_{2i} = \mathsf{HE}_{pk}(r_iy_if(y_i))$,$c_{2i}$ 需要重加密，即 c_{1i}, c_{2i} 不能使用相同的随机数加密，否则 S 获得 $c_{1i} = \mathsf{HE}_{pk}(r_if(y_i)), c_{2i} = \mathsf{HE}_{pk}(r_iy_if(y_i))$，两者相差 y_i，S 可以暴力测试出 y_i 的信息（当 Y 的空间较小时）；

该方案和 Frikken-07 相比基本没什么优势，毕竟第一轮消息依旧和 X 相关（多项式次数）；但是因为多项式计算转换到了 S，所以有更好的优化技术——FHE；使用 FHE 加密，R 可以只加密一个元素 $\mathsf{FHE}(y_i)$，优化了第一轮的通信量；

6 Two-party FHE-based PSU

$S(x_i \in X, i \in [n])$		$R(y_i \in Y)$
$f(x) = a_1 + a_2x + \cdots + a_nx^{n-1}$	$\xleftarrow{\mathsf{FHE}_{pk}(y_i)}$	
	$\xrightarrow{c_i=\mathsf{FHE}_{pk}(f(y_i))}$	$r_i \leftarrow \mathbb{Z}_q, c_{1i} = \mathsf{FHE}_{pk}(r_if(y_i))$
		$c_{2i} = \mathsf{FHE}_{pk}(r_iy_if(y_i))$
	$\xleftarrow{c_{1i},c_{2i},d_{R1i},d_{R2i}}$	$d_{R1i} = \mathsf{Dec}_{sk_R}(c_{1i}), d_{R2i} = \mathsf{Dec}_{sk_R}(c_{2i})$
$d_{S1i} = \mathsf{Dec}_{sk_S}(c_{1i}), d_{S2i} = \mathsf{Dec}_{sk_S}(c_{2i})$		
$m_{1i} = \mathsf{Combine}(d_{S1i}, d_{R1i})$		
$m_{2i} = \mathsf{Combine}(d_{S2i}, d_{R2i})$		
If $m_{1i} = m_{2i} = 0$, $y_i \in X$		
If $m_{1i} \neq 0, m_{2i} \neq 0$, $y_i = m_{2i}/m_{1i} \notin X$		

关键点：$c_{1i}=\mathsf{FHE}_{pk}(r_i f(y_i)), c_{2i}=\mathsf{FHE}_{pk}(r_i y_i f(y_i))$,$c_{2i}$ 需要重加密，即 c_{1i}, c_{2i} 不能使用相同的随机数加密，否则 S 获得 $c_{1i}=\mathsf{FHE}_{pk}(r_i f(y_i)), c_{2i}=\mathsf{FHE}_{pk}(r_i y_i f(y_i))$，两者相差 y_i，S 可以暴力测试出 y_i 的信息（当 Y 的空间较小时）；

7 Conclusions

Frikken-07 方案的通信轮数 2 轮，通信复杂度 $|X|+2|Y|$

TP-FHE 方案的通信轮数 3 轮，通信复杂度 $|Y|+4|Y|$，所有的通信量和 Y 相关，如果 Y 很小的话 (unbalance 场景下)，通信效率较高。

如果不考虑两方 FHE，FHE 只能判定 y 中元素是否在 X 中，后续则需要使用 OT 提取 Y 中的元素；判定部分的通信量为 $O(3|Y|)$;

参考文献