Midterm Report

牛午甲 PB20111656
潘云石 PB20111657
石磊鑫 PB20111658
孙霄鹃 PB20111659
陈　昊 PB20051077

Abstract

经过一段时间对隐私集合求交（PSI）的学习，我们已经基本掌握了PSI中的问题模型和现有的几种协议。并且我们对比了这些协议的优缺点。另外，我们也学习了PSI中的效率与授权问题，对于Practical Private Set Intersection Protocols with Linear Complexity一文中的高效PSI算法，我们准备研究将其拓展为authorized PSI(APSI)算法。

Progress

> Research Problems
>
> - 目前大多数对于 PSI 的研究都是基于半诚实模型。在 Malicious Model 中，协议需要使用额外的手段来降低被攻击的可能，因此 Malicious Model 下安全的协议的复杂程度和开销往往远大于半诚实模型下安全的协议，有没有可能基于其他出发点，设计性能接近半诚实方案的 Malicious PSI 模型?
> - 目前非平衡隐私集合求交的应用场景有很多，如某些 APP 的联系人寻找，而目前调研到的文章没有提到对于非平衡隐私集合求交的特别处理，即一个集合很小，一个集合很大时的情况，当前的非平衡PSI协议相较于平衡的PSI协议性能优势并不是很明显，是否可以通过特殊的办法使得非平衡隐私集合求交的效率显著高于平衡的隐私集合求交?
> - 除了目前工作的理论支撑外，有没有可能存在其他理论方法可以用来解决 PSI 问题?

对于我们在 Proposal 中 Research Problems 部分提出的问题，我们有针对性地进行了进一步调研，试图从现有的工作中获得一些启发。在此过程中我们了解了多种技术的优势和不足，并做了比较：(不引起混淆时记 $N = N_x = N_y$)

分类	协议	计算复杂度	通信复杂度	优势	不足	启发
安全受限	Naive PSI	$O(N^2)$	$O(N)$单向通信一次	计算开销小，速度快	很容易被攻击	无
基于公钥体系(DH)	APSI	$O(N^2)$ 有较多幂运算	$O(N)$双向通信一次	一定程度上抵御malicious party攻击	参与方计算开销差异明显， 其中一个参与方需要进行预计算	可以考虑进一步减小开销，作为Problem 1的解决方案
基于garbled circuit	SCS	$O(N \log N)$	$O(N \log N)$	通用协议，易于模块化设计	消耗带宽大	在保障安全的前提下通过排序可以降低求交集的复杂度
基于garbled circuit	GMW+SCS	$O(N \log N)$ 在OT上进行优化， 降低了计算开销	$O(N \log N)$	通用协议，易于模块化设计	消耗带宽大	在保障安全的前提下适量减小随机性可以提高效率
基于 OT	Random GBF-based	$O(N)$	$O(N)$ 双向通信两次以上	设计了高效数据结构， 极大提高了空间效率和查询时间	有一定的误识别率； 集合元素修改较困难	可以借助GBF快速判断元素是否在集合中

分类	协议	计算复杂度	通信复杂度	优势	不足	启发
基于哈希	Cuckoo hashing-based	$O(N \log\log N)$	$O(N)$ 双向通信两次以上	可以减少必须计算的比较次数	存在哈希失败的概率	可以将Bucket Sorting、Rabin Karp算法的思想用在提升比较效率上
基于 OT	BaRK-OPRF	$O(N^2)$	$O(N)$ 双向通信两次以上	极大优化OT过程：使用 k 次 1-out-of-2 OT 实现实际上 $m(>> k)$ 次 1-out-of-2 OT，并且没有消息长度限制	接收方求交集的开销较大，成为性能瓶颈	可以将编码和加密同时应用于OT
基于FHE	FHE-PSI	$O(N^2)$	$O(min\{N_x, N_y\})$ 双向通信一次	利用FHE大幅度降低通信开销	由于需要FHE，本地计算开销巨大	可以作为Problem 2的解决方案的出发点

Challenges and Obstacles

1. 对于我们在 Proposal 中 Methodologies 部分提到的几种理论上或许可行的 PSI 方法，理论上是否存在安全隐患是一个很难判断的问题。我们将从现有的其他协议下的攻击方法出发，尝试寻找攻击的角度。
2. 目前没有便于我们测试PSI算法的线上平台，我们或许需要自己写网络应用。

Left Workloads

1. 对于我们在 Proposal 中 Methodologies 部分提到的几种理论上或许可行的 PSI 方法，目前还没有进行进一步的思考与完善。我们将在后续工作中尽可能地提出进一步的思路，甚至给出初始协议和简单的安全性说明。
2. 经过前期调研，我们打算在非对称 PSI 场景下做一些实验上的工作。我们准备先复现Practical Private Set Intersection Protocols with Linear Complexity一文中的4种PSI算法（其中2种是APSI）。论文最后运用数据预处理等技术给出了一个高效的PSI算法，我们准备进行研究将其拓展为APSI版本。这也是作者提及的Drawback之一。

   > Drawbacks: although very efficient, this PSI protocol has some issues. First, it is unclear how to convert it into an APSI version.