计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.20220471
59(10):2286-2298,2022

高效且恶意安全的三方小集合隐私交集计算协议

张 蕾[1] 贺崇德[1] 魏立斐[2]
[1](上海海洋大学信息学院 上海 201306)
[2](上海海事大学信息工程学院 上海 201306)
(Lzhang@shou.edu.cn)

Efficient and Malicious Secure Three-Party Private Set Intersection Computation Protocols for Small Sets

Zhang Lei[1], He Chongde[1], and Wei Lifei[2]
[1](*College of Information Technology, Shanghai Ocean University, Shanghai* 201306)
[2](*College of Information Engineering, Shanghai Maritime University, Shanghai* 201306)

Abstract Private set intersection (PSI) allows participants who hold private sets to securely obtain the set intersection without revealing information about any elements other than the intersection. Most of the existing two-party/multi-party PSI protocols are based on the oblivious transfer (OT) protocol, which not only has high efficiency, but also brings huge cost of communication. However, expanding network bandwidth is very expensive or even infeasible in many scenarios, and there are few computationally efficient multi-party PSI protocols that do not rely on OT protocols. In this paper, three-party private set intersection computing protocols are constructed based on one round key agreement. The protocols are proved secure assuming the collusion attack of any two parties in the semi-honest model and malicious model, respectively. Through experimental simulation, in the large set scenario, compared with the existing OT-based multi-party PSI protocol, the three-party private set intersection computation protocols have the optimal number of communication rounds, and the amounts of communication is reduced by 89%~98%. In small set scenarios (500 items or less), compared with similar PSI protocols for weak communication networks, the protocols in our paper have optimal runtime and communication load, especially, and they get 10~25 times faster than PSI protocol relying on homomorphic encryption.

Key words private set intersection(PSI); key agreement; malicious adversaries; collusion resistant attack; small sets scenarios

摘 要 隐私集合交集(private set intersection, PSI)允许持有私有集合的参与方安全地获得集合的交集,而不会泄露除交集之外任何元素的信息.现有的两方/多方 PSI 协议大多基于不经意传输(oblivious transfer, OT)协议,具有很高计算效率的同时,也带来了巨大通信开销.在很多场景中,扩展网络带宽是

收稿日期:2022-06-10;**修回日期**:2022-08-10
基金项目:国家自然科学基金项目(61972241);上海市自然科学基金项目(22ZR1427100,18ZR1417300);上海市高可信计算重点实验室开放课题(OP202102);上海市青年科技英才扬帆计划(21YF1417000);上海海洋大学骆肇尧大学生科技创新基金项目
This work was supported by the National Natural Science Foundation of China (61972241), the Natural Science Foundation of Shanghai (22ZR1427100,18ZR1417300); The Open Project of Shanghai Key Laboratory of Trustworthy Computing (OP202102); Shanghai Sailing Program(21YF1417000); Luo Zhaorao College Student Science and Technology Innovation Fund of Shanghai Ocean University.
通信作者:魏立斐(lfwei@shmtu.edu.cn)

非常昂贵甚至不可行的,而目前不依赖于 OT 设计且计算高效的多方 PSI 协议仍然较少.基于一轮密钥协商构造了三方参与的 PSI 计算协议,分别在半诚实模型和恶意安全性模型下,证明了协议的安全性且允许任意两方的合谋攻击.通过实验仿真,在大集合场景,相比现有基于 OT 的多方 PSI 协议,所构造的协议具有最优的通信轮数且通信量降低了 89%～98%;在小集合场景(500 个元素或更少),相比适用弱通信网络的同类 PSI 协议,具有最优运行时间和通信负载,比依赖于同态加密的 PSI 协议快 10～25 倍.

关键词　隐私集合交集;密钥协商;恶意敌手;抗合谋攻击;小集合场景

中图法分类号　TP309

隐私集合交集(private set intersection, PSI)技术是安全多方计算技术的重要应用之一.经典的 PSI 协议允许 2 个参与方各自持有自己的私有集合,在协议结束时两方或其中一方作为接收者,获得两方集合的交集,且不会泄露除交集之外的任何元素信息[1-2].作为重要的密码学工具,PSI 也被广泛应用于人工智能和数据挖掘的安全领域,如隐私保护数据挖掘[3-4]、私有通讯录查找[5]、新冠接触者追踪[6]以及衡量在线广告转化率[7]等.在数据共享的时代背景下,多方参与 PSI 的场景需求更加广泛,例如在社交软件的隐私联系人查找功能中,可查找多个用户的共同好友即属于多方参与 PSI 的应用场景.

目前大多数高效的 PSI 方案都是基于不经意传输(oblivious transfer, OT)协议[8]构建,得益于高效的 OT 扩展技术,各方可以通过少量的公钥操作生成大量的 OT 协议实例,使得基于 OT 的 PSI 协议所需要的公钥操作数量仅与安全参数有关,而与集合大小无关,计算成本较低,从而高效地构造 PSI 协议[9-15],但其通常具有一定的固定成本,往往仅适合大集合(2^{10}～2^{20}个元素)场景,而在小集合(500 个元素或者更少)场景下优势不够明显.此外,虽然基于 OT 的协议计算效率较高,但同时带来了巨大的通信成本,而在某些实际场景下通信成本比计算成本更重要[16].

基于密钥协商构造的 PSI 协议[17-20]一般通信量较低,在弱通信场景中具有较大优势.例如,采用目前最有效的 1-out-of-2 OT[21]构建 PSI,128 个基本 OT 需要花费 384 个群元素进行通信以及 640 次指数运算,其开销比集合大小为 200 的基于 Diffie-Hellman 密钥协商的 PSI 还要昂贵[16].从实用成本的角度,在网络中添加 CPU 比扩大网络容量要便宜的多,因此,在谷歌内部部署 PSI 功能时选择了基于 Diffie-Hellman 密钥协商的 PSI[16].此外,基于 RSA 和全同态加密的协议也具有很低的通信成本[22-26],常被用于弱通信场景中,但其采用大量繁重的公钥操作,产生了巨大的计算成本,导致非常低的运算效率.

小集合交集计算是 PSI 协议的一个典型场景,具有广泛的应用.例如,为了增强了苹果手机的隔空投送功能,文献[27]通过对用户的整个地址薄(几千项)和另一个用户的个人标识符(电话号码或电子邮箱,可能是 10 项)进行 PSI.又如,各方可能希望利用可用日历时间的 PSI 来安排在线会议时间,即各自可用时间集合(按小时划分,此时集合规模约为 360 小时[20])的交集.对于此类输入大小的集合,基于密钥协商的 PSI 是计算成本最低的.Rosulek 等人[20]在 CCS 2021 上采用 Diffie-Hellman 密钥协商和多项式插值技术,在小集合情况下实现了迄今为止最快的 PSI 方案.然而,文献[20,27]所述的基于密钥协商的方案都仅适用于具有 2 个参与方的场景.PSI 协议的隐私性要求除交集之外的任何信息都无法被泄露,而两方协议直接扩展到多方将不可避免地泄露交集之外的两两相交的部分,导致两方协议无法直接扩展到多方.

综上,本文提出了一个基于密钥协商的三方恶意安全 PSI 协议,能抵抗任意 2 个恶意参与方合谋,实现了现有多方 PSI 中最低的通信量;特别地,在小集合场景下,保持了较高的运算效率.该协议非常适合三方小集合场景,例如为了签订合同,投资方、劳务方和中介方希望利用 1 个月内的可用时间安排会议,需要 3 个参与方利用各自可用时间的集合,考虑存在合谋的情况下,不泄露各方集合信息时求出交集.本文的主要贡献有 2 个方面:

1) 提出了一个基于密钥协商的三方 PSI 协议,实现了半诚实的安全性,允许任意 2 个参与方合谋,并在此基础上提出了恶意安全的三方 PSI 协议.利用模拟范式,证明了协议在半诚实和恶意安全模型下合谋时的安全性.

2) 通过实验仿真,在大集合场景,相比现有基于 OT 的多方 PSI 协议,本文协议具有最优的通信

轮数,通信量降低了89%～98%;在小集合场景,相比适用弱通信网络的同类PSI协议,具有最优的运行时间和通信负载,其中运算速度比依赖于同态加密的PSI协议快10～25倍.

1 相关工作

PSI作为安全多方计算的热点问题已经得到了快速的发展.经典的PSI协议包含2个参与方,已经达到非常高的效率.最早的PSI协议采用朴素哈希的方式,即先对集合元素求哈希,并通过对哈希值的对比得出交集,这种方案是十分高效的,但容易受到碰撞攻击.为了解决碰撞攻击的问题,需要采用安全对比的方法.对于持有 m 个元素的集合的双方,为了求出交集元素,最坏的情况需要进行 $O(m^2)$ 次比较.通过将元素映射到长为 n 的布隆过滤器,比较次数可减少到 $O(n\,\mathrm{lb}\,n)$.随着PSI技术的成熟,通过布谷鸟哈希、不经意伪随机函数和高效的OT扩展等技术,一些PSI协议实现了 $O(n)$ 的计算和通信复杂度.

最新对抗半诚实敌手的两方PSI协议来自文献[9].文献[9]设计了一种基于OT技术的安全字符串相等性测试协议,仅使用OT[8]、哈希函数、对称密钥加密操作和按位操作来构建协议,因此计算效率很高.文献[10]基于OT扩展和多项式编码技术实现,除基于昂贵的RSA和FHE的协议之外,该协议在已有的两方PSI中具有最低的通信.文献[11]提出了一种高效的多点不经意伪随机函数(oblivious pseudo-random function, OPRF),实现了计算和通信之间更好的平衡,在具有中等带宽的网络中达到了所有已知协议中最高的运行效率.最新恶意安全两方PSI协议是文献[28]和文献[29],它们分别基于高效的OT扩展和向量OLE[30].当集合比较大(例如 $n>2^{20}$)时,文献[29]非常高效,但它具有较高的固定成本,这使得它在较小的集合场景中效率较低.

随着隐私集合交集技术的成熟及其在实际应用中的普及,2个参与方已经不能满足应用需求,多参与方的场景更加广泛,但目前仅有少数方案适用于此类场景.第1个高效的多方PSI协议由Kolesnikov等人[31]提出,该协议同样使用OT扩展实现,他们为协议设计了2个版本,分别实现了半诚实及增强半诚实(可能在执行开始前改变攻陷参与方的输入)的安全性.Efraim等人[32]巧妙地结合了来自半诚实安全的多方PSI[33]和恶意两方PSI[34]的结果,提出了第1个恶意安全的多方PSI,但该协议需要传输混乱布隆过滤器(garbled Bloom filter, GBF)进行通信,这带来了巨大的通信负担.Nevo等人[35]首先使用高效的不经意键值存储(oblivious key-value stores, OKVS)技术[36]和高效不可信云辅助两方PSI实现了迄今为止最快的恶意多方PSI协议,但该协议在恶意参与者合谋的情况下是不安全的,会泄露其他方的集合元素信息.利用不经意零共享技术和不经意可编程伪随机函数(oblivious programmable pseudorandom function, OPPRF),Nevo等人[35]在第2个协议中实现了允许任意 t 个参与方进行合谋的恶意PSI协议,实现了已有恶意协议中最好的计算和通信效率.

基于密钥协商构建PSI协议是隐私集合交集计算的经典思路,也是本文工作的重点.最早的基于密钥协商的PSI协议可以追溯到1999年提出的文献[17],并实现了半诚实的安全性,但由于该协议完全按照Diffie-Hellman协议设计,导致其在半诚实安全下变体的设计空间非常有限.之后的研究中,基于Diffie-Hellman的PSI协议在恶意敌手存在情况下的安全性得到增强.文献[18-19]提出了高效且恶意安全的PSI方案,实现了线性的通信复杂度和线性的计算复杂度.最新的文献[20]利用多项式插值将参与方集合元素映射到Diffie-Hellman密钥协商的密钥空间,通过对比输出密钥得出交集,分别实现了半诚实和恶意安全性,并极大地提高了基于密钥协商的PSI的效率,尤其是在小集合的情况下,该方案提出的恶意安全协议甚至比同类半诚实安全协议的运算速度快,同时通信成本降低了40%.但上述基于密钥协商的PSI协议都仅适用于两方,且无法直接扩展到多个参与方的场景.

2 预备知识

2.1 安全模型

安全多方计算的安全模型可以分为半诚实模型和恶意模型2种.在半诚实模型下,敌手完全遵循协议的执行过程,但可能会记录协议执行过程中的所有数据,并试图从协议执行过程数据中获取额外信息;在恶意模型中敌手不仅可以通过协议过程的数据推测敏感信息,还可以不遵循协议规范,拒绝参与协议、修改隐私的输入集合信息或者提前终止协议的执行等.本文提出的协议分别在半诚实模型和恶

意模型下可证安全而不泄露任何信息,且允许任意2个参与方的合谋.

针对安全多方计算的协议普遍采用模拟范例进行证明,将理想状态下引入可信第三方的安全多方计算协议的理想协议与真实协议进行对比,如真实协议的视图与理想协议的视图是不可区分,则真实协议未泄露更多信息,进而证明协议是安全的.在半诚实模型中,模拟器仅需根据协议规则模拟敌手视图,并证明与真实协议的视图是不可区分.在恶意模型中,需要进一步考虑敌手的恶意行为,对恶意参与方进行输入提取,并考虑其对诚实参与方输出的影响严格进行模拟,最终证明模拟器与真实协议的视图不可区分.

本文协议包含3个参与方,可以抵抗任意两方的合谋攻击.在三方协议中,需要对任意两方合谋进行模拟,证明模拟器视图与真实协议视图计算不可区分.

定义1. 半诚实模型安全性.设 f 是理想三方协议,用 I 表示任意参与方子集.I 中参与方在执行输入为 (x,y,z) 三元组的协议 π 时,其视图表示为 $VIEW_I^{\pi}(x,y,z,n)=(\overline{X},r_1^I,\cdots,r_t^I,m_1^I,\cdots,m_k^I)$,其中 n 为安全参数,$\overline{X}$ 表示 I 中各参与方的输入值列表,r_j^I 表示参与方集合 I 在执行过程中产生的第 j 个随机数,m_j^I 表示 I 收到的第 j 个消息.如果存在使用概率多项式时间算法的模拟器 S,使得对于任意的 I 均有下式成立:

$$S_I(\overline{X},f(x,y,z,n))\overset{c}{\equiv}VIEW_I^{\pi}(x,y,z,n),$$

则称协议 π 安全地计算了 f,其中 $\overset{c}{\equiv}$ 表示计算不可区分.

定义2. 恶意模型安全性.设 f 是理想状态下的三方协议,π 为真实协议.如果对于现实模型中所有使用概率多项式时间算法的敌手 A 都存在一个理想模型中使用概率多项式时间算法的模拟器 S,使得

$$IDEAL_S^f(x,y,z,n)\overset{c}{\equiv}REAL_A^{\pi}(x,y,z,n),$$

即敌手 A 在真实协议中得到的信息与理想模型得到的信息是不可区分,则称协议 π 在恶意敌手存在的情况下安全地计算了 f,其中 x,y,z 为参与方的输入,n 为安全参数.

2.2　三方PSI的理想功能

3个参与方各自拥有自己的私有集合,通过联合执行三方PSI协议,接收方获得3个参与方集合交集元素的信息,除此之外各参与方无法获取任何额外信息.另外,敌手会设置状态 $abort$,若 $abort=1$,则协议会中止.特别地,在半诚实模型中,敌手总是设置 $abort=0$.图1描述了三方PSI的理想功能,其中 $[n]$ 表示整数集合 $\{1,2,\cdots,n\}$.

参数:3个参与方拥有的集合大小 n.
功能:参与方 P_1,P_2,P_3 输入各自的私有集合 $A^i=\{a_1^i,a_2^i,\cdots,a_n^i\}\subset\{0,1\}^*$ 以及由敌手产生的 $abort\in\{0,1\}$.若 $abort=0$,输出 $\bigcap_{i\in[3]}A^i$ 给接收方 P_3;否则输出 $\perp$ 给 P_3.

Fig. 1　Ideal functionality of three-party PSI
图1　三方PSI的理想功能

2.3　理想置换

在理想置换模型中,各方可以访问 $\{0,1\}^n$ 上的随机置换 Π 及其逆置换 Π^{-1},在安全性证明中,模拟器可以观察理想置换 Π,Π^{-1} 上所有查询并编码响应,为模拟敌手的视图提供帮助.目前,理想置换已被用于实现混乱电路和OT协议中哈希函数[37-38].理想置换也可以用于设计安全的PSI协议,文献[20]利用理想置换分别实现了半诚实和恶意敌手的安全性.

2.4　双线性配对

双线性配对广泛应用于密码方案的设计[39].设 G 是素数 q 阶加法循环群,G_T 是 q 阶乘法循环群.映射 $e:G\times G\to G_T$ 满足下列性质,则被称为一个双线性配对映射:

1) 双线性.对于任意的 $P,Q\in G,a,b\in\mathbb{Z}_p^*$,则有 $e(aP,bQ)=e(P,Q)^{ab}$.

2) 非退化性.存在 $P,Q\in G$,使得 $e(P,Q)\neq 1_{G_T}$,其中 1_{G_T} 是 G_T 中的单位元.

3) 可计算性.对于任意的 $P,Q\in G$,存在有效的多项式时间算法可以计算 $e(P,Q)$.

本文方案的安全性主要基于判定性双线性(decision bilinear Diffie-Hellman, DBDH)假设:对于 $P\in G$,给定 (P,aP,bP,cP) 和元素 $h\in G_T$,判断 $e(P,P)^{abc}$ 和 h 是计算不可区分的,即对于任意的多项式时间算法 F 及任意的 n,均有:

$$Pr[F(P,aP,bP,cP,e(P,P)^{abc})=1]-Pr[F(P,aP,bP,cP,h)=1]\leqslant\varepsilon(n),$$

其中 $\varepsilon(n)$ 是参数为 n 的可忽略函数.

2.5　三方密钥协商协议

本文设计的三方PSI协议是基于文献[40]提出的三方Diffie-Hellman密钥协商协议.文献[40]的协议仅需要一轮通信即可构建一个共享的密钥.三方密钥协商协议过程如图2所示.该协议包含3个

参与方 A,B,C,给定双线性配对 $e:G\times G\rightarrow G_T$,其中 G 是素数 q 阶加法循环群,G_T 是 q 阶乘法循环群.P 是 G 的一个生成元,A,B,C 各随机选择 a,b,$c\in\mathbb{Z}_p^*$,分别广播 aP,bP,cP.根据双线性配对性质,三方均可以计算共享密钥 $K=e(bP,cP)^a=e(aP,cP)^b=e(aP,bP)^c=e(P,P)^{abc}$.

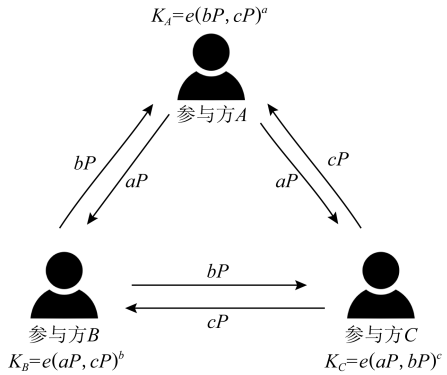

Fig. 2 Three-party key agreement protocol

图 2 三方密钥协商协议

3 三方隐私集合交集计算协议

本节首先介绍了第一个半诚实版本的协议 1(PSI-s),主要思想来源于三方密钥协商[40],仅需两轮通信.协议要求各参与方是半诚实的,同时允许任意两方合谋.通过改进半诚实版本的协议,在协议 2(PSI-m)中达到了恶意攻击安全.在协议描述中,用 K 表示密钥协商中输出密钥空间,κ 表示计算安全参数,λ 表示统计安全参数.

3.1 半诚实三方隐私集合交集协议(PSI-s)

3.1.1 基本协议

基于密钥协商的三方 PSI 协议的主要想法是将参与方的元素与输出密钥用插值多项式联系起来:若元素在集合的交集中,则输出相同的密钥,否则将输出不同的密钥,这个过程不会泄露任何信息.

协议模型结构如图 3 所示.3 个参与方 A,B,C,其中参与方 C 作为接收者,在协议执行结束之后获得三方集合交集的元素信息.协议将集合元素映射到密钥协商后的公共密钥空间上,若三方获得的公共密钥相同,则可判断对应的元素在交集中,否则该元素不在交集中.事实上,将元素映射到 3 个参与方的密钥空间是不必要的,本节中提出的协议仅仅将元素映射到参与方 B,C 的密钥空间,这大大降低了协议所需的计算量和通信量.

半诚实三方 PSI 的完整协议在协议 1 中给出.参与方 A 通过插值多项式,将所有 $x_i\in X$ 与用于生成共享密钥的参数 a_iP 进行关联,并通过理想置换 $\Pi^{-1}(a_iP)$使得生成的多项式与随机选取的同阶多项式不可区分,从而保证了集合 X 中元素的隐私性.参与方 B 和 C 通过对来自 A 的多项式求响应,并最终求出共享密钥,分别隐含了 $X\cap Y$ 和 $X\cap Z$ 的信息.最终 C 通过对比共享密钥,即可得出三方集合交集,不会泄露除交集外的任何元素信息.

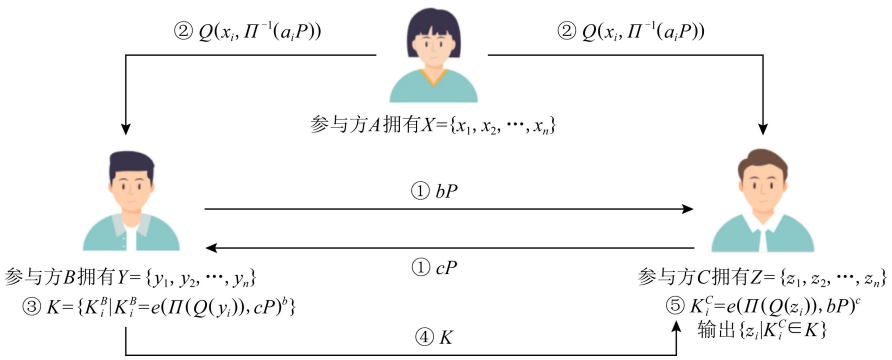

Fig. 3 Semi-honest three-party private set intersection computation protocol

图 3 半诚实三方隐私集合交集计算协议

协议 1. 基于密钥协商的半诚实三方 PSI(PSI-s).

参数:3 个参与方 A,B,C;有限域 F,循环群 G 和 G_T,P 是 G 的生成元;理想置换 Π,Π^{-1}:$F\rightarrow F$;双线性配对 $e:G\times G\rightarrow G_T$;随机密钥空间 $|K|\geqslant 2^{\lambda+2\,\mathrm{lb}\,n}$.

输入:参与方 A,B,C 分别输入集合 X,Y,Z;

输出:接收方 C 输出集合 $X\cap Y\cap Z$.

协议过程如下:

1) B 随机选择 $b\in\mathbb{Z}_p^*$ 计算 bP 发送给 C,同时 C 随机选择 $c\in\mathbb{Z}_p^*$ 计算 cP 发送给 B.

2) A 随机选择 n 个随机值 $\{a_1,a_2,\cdots,a_n\}\in\mathbb{Z}_p^*$,插值多项式 $Q(x_i,\Pi^{-1}(a_iP))$并发送给 B,C.

3) B 计算输出密钥.B 对每一个 $y_i \in Y$,对来自参与方 A 的多项式求值 $Q(y_i)$,并计算每个响应对应的共享密钥值 $K=\{K_i^B \mid K_i^B=e(\Pi(Q(y_i)),cP)^b\}$,将其乱序发送给 C.

4) C 计算输出密钥.C 和 B 做同样的操作,即对每一个 $z_i \in Z$,对多项式求值 $Q(z_i)$,并计算每个响应对应的共享密钥值 $K_i^C=e(\Pi(Q(z_i)),bP)^c$.

5) 计算交集.参与方 C 判断 K_i^C 是否在 K_B 中,若 $K_i^C \in K$,则输出 x_i^C.

3.1.2　正确性分析

参与方 A 插值多项式 Q 使得 $Q(x_i)=\Pi^{-1}(a_iP)$,参与方 B 用自己的集合元素在多项式上求响应并进行理想映射得到 $\Pi(Q(y_i))$.若参与方 A 与 B 拥有相同的集合元素 $x_i=y_i$,则 $\Pi(Q(y_i))=a_iP$,然后利用双线性配对计算输出密钥集合 $K=\{K_i^B \mid K_i^B=e(\Pi(Q(y_i)),cP)^b\}$,此时 $K_i^B=e(a_iP,cP)^b$,否则 K_i^B 将是一个随机值.同理,参与方 C 用自己的集合元素在多项式上计算并进行理想置换得到 $\Pi(Q(z_i))$,若参与方 A 与 C 拥有相同的集合元素 $x_i=z_i$,则 $\Pi(Q(z_i))=a_iP$,当参与方 C 利用双线性配对计算输出密钥 $K_i^C=e(\Pi(Q(z_i)),bP)^c$,此时 $K_i^C=e(a_iP,bP)^c$,否则将是一个随机值.

因此,对于交集中的元素 $x_i=y_i=z_i$,有:

$$K_i^C=e(\Pi(Q(z_i)),bP)^c=e(a_iP,bP)^c=e(P,P)^{a_ibc}=e(a_iP,cP)^b=e(\Pi(Q(y_i)),cP)^b=K_i^B.$$

因此,满足 $K_i^C \in K$ 的元素,即交集 $X \cap Y \cap Z$ 的元素.

假设 $x_i \notin X \cap Y \cap Z$,只有当 x_i 对应的随机密钥 $K_i^C=K_i^B$ 时,接收方才会产生不正确的输出.对于这个特殊的 x_i,当随机密钥的随机空间足够大时,这个事件发生的概率是可忽略的.设随机密钥空间为 K,则该事件发生的概率为 $n/|K|$,在最多 n 个这样的值时,发生的总概率为 $n^2/|K|$,由于协议的随机密钥空间 $|K| \geqslant 2^{\lambda+2\mathrm{lb}n}$,故发生错误的概率是 $2^{-\lambda}$,即可忽略的.

3.1.3　安全性证明

定理 1. 假设 Π,Π^{-1} 是理想置换,协议 1 在半诚实模型下安全地实现了三方隐私集合交集计算,其安全性归约于判定性双线性问题的困难性假设.

证明. 在本文的协议中,任意 2 个参与方合谋即为最大的合谋攻击,因此只需要证明协议对任意 2 个参与者合谋时是安全的,则对于任意单个半诚实参与方也是安全的.分 3 种情况分别讨论.

1) 半诚实的参与方 A 与 B 合谋情况

在参与方 A,B 合谋的情况下,敌手从诚实参与方 C 接收到的唯一消息是 cP,由于 c 是在 $\mathbb{Z}_p^*$ 中随机选取的,cP 和 G 中随机值是不可区分的.因此,在 A,B 合谋的情况下,A,B 未获得任何有用信息.

2) 半诚实的参与方 B 与 C 合谋情况

在参与方 B,C 合谋的情况下,敌手获得的消息是来自诚实参与方 A 的多项式 $Q(\cdot)$,因 Π 是理想置换,每一个 $\Pi^{-1}(a_iP)$ 值与随机值是不可区分的,因此 $Q(\cdot)$ 是一个输出随机均匀的多项式而与输入 x 无关,$Q(\cdot)$ 与随机选取的同阶多项式是无法区分的,B,C 未获得任何有用信息.

3) 半诚实的参与方 A 与 C 合谋情况

设计以下模拟器 S 来证明它和真实协议是不可区分的.定义混合序列 $hybrid_h$:步骤①S 按照协议规则诚实地发送 bP 给参与方 C;步骤②对于 $z_i \in (X \cap Y \cap Z) \cup \{z_i \mid i \geqslant h\}$,模拟器计算 $k_i=e(\Pi(Q(z_i)),bP)^c$,其中 Π 由参与方 A 决定;步骤③对于所有其他的 $z_i \in Z$,选择随机值作为 k_i 的输出;步骤④S 生成集合 K.

此时,$hybrid_0$ 即为真实协议,而 $hybrid_n$ 即为设计的模拟器.在 $hybrid_n$ 中只利用了交集 $X \cap Y \cap Z$ 中的元素生成信息,所以不会泄露不在交集中元素的任何信息.

为了证明 $hybrid_h$ 和 $hybrid_{h+1}$ 是不可区分的,修改上述混合序列步骤③为:若 $z_h \notin X \cap Y \cap Z$,模拟器计算 $k_h=k^*$;对于所有其他的 $z_i \in Z$,选择随机值作为 k_i 的输出.

此时,上述混合序列在 k^* 被赋予不同的值时分别对应于 $hybrid_h$ 和 $hybrid_{h+1}$.在 $hybrid_h$ 中,k^* 被计算为 $k^*=e(\Pi(Q(z_i)),bP)^c$,而在 $hybrid_{h+1}$ 中 k^* 被赋予随机值.根据参数选取的随机性和双线性配对的性质易知,$hybrid_h$ 和 $hybrid_{h+1}$ 是不可区分的,从而证明了模拟器和真实协议也是不可区分的.

综合 1)~3)知,协议 1 在半诚实模型下安全地实现了基于密钥协商的三方隐私集合交集计算且能够抵抗合谋攻击.　　证毕.

3.2　恶意安全三方隐私集合交集协议(PSI-m)

3.2.1　基本协议

本节将展示恶意安全的基于密钥协商的三方 PSI 协议.恶意安全三方 PSI 的完整协议在协议 2 中给出.为了对抗恶意敌手,让参与方 A 插值多项式 Q,使得 $Q(H(x_i))=\Pi^{-1}(a_iP)$,其中 Π 是理想置换,

H 是抗碰撞哈希函数,将参与方 B 的输出 K 替换为 $K=\{K_i^B \mid K_i^B=H_2(y_i,k_i)\}$.

在半诚实版本的基础协议中存在以下安全威胁:若恶意参与方 A 插值多项式使得 $Q_1(b)=\Pi^{-1}(a_bP)$,其中,b 是参与方 B 的元素,a_b 是与 b 对应的随机数.参与方 A 将此多项式发送给 B.接下来 A 插值多项式使得 $Q_2(c)=\Pi^{-1}(a_bP)$,其中 c 是参与方 C 的元素.根据协议过程,该行为将导致协议输出错误的结果(元素 c 并不在交集,但它将会被输出).本文在恶意版本的协议 2 中将参与方 B 的输出 K 替换为 $K=\{K_i^B \mid K_i^B=H_2(y_i,k_i)\}$避免了这种攻击.在安全性证明中,$H_1$ 和 H_2 作为随机预言机来帮助模拟器进行输入提取.

协议 2. 基于密钥协商的恶意三方 PSI(PSI-m).

参数:3 个参与方 A,B,C;有限域 F,循环群 G 和 G_T,P 是 G 的生成元;理想置换 $\Pi,\Pi^{-1}:F\to F$;双线性配对 $e:G\times G\to G_T$;抗碰撞的哈希函数 $H_1:\{0,1\}^*\to F$,$H_2:\{0,1\}^*\times G_T\to\{0,1\}^{2\kappa}$;随机密钥空间 $|K|\geqslant 2^\kappa$.

输入:参与方 A,B,C 分别输入集合 X,Y,Z;

输出:接收方 C 输出集合 $X\cap Y\cap Z$.

协议过程如下:

1) B 随机选择 $b\in\mathbb{Z}_p^*$ 计算 bP 发送给 C,同时 C 随机选择 $c\in\mathbb{Z}_p^*$ 计算 cP 发送给 B.

2) A 随机选择 n 个随机值 $\{a_1,a_2,\cdots,a_n\}\in\mathbb{Z}_p^*$,并对每一个 $x_i\in X$ 求哈希值 $H_1(x_i)$. 随后插值多项式 $Q(H_1(x_i),\Pi^{-1}(a_iP))$发送给 B,C.

3) B,C 接收多项式,若多项式阶数小于 1,则协议终止.

4) B 计算输出密钥.B 对每一个 $y_i\in Y$ 求哈希值 $H_1(y_i)$,然后用 $H_1(y_i)$对来自参与方 A 的多项式 $Q(\cdot)$求值,并计算对应的共享密钥值 $k_i=e(\Pi(Q(H_1(y_i))),cP)^b$.

5) B 对每个 $y_i\in Y$ 和相应的 k_i 联合求哈希值得到 $K=\{K_i^B \mid K_i^B=H_2(y_i,k_i)\}$,将其乱序发送给 C.

6) C 计算输出密钥.C 对每一个 $z_i\in Z$ 求哈希值 $H_1(z_i)$,然后用 $H_1(z_i)$对来自参与方 A 的多项式 $Q(\cdot)$ 求值,并计算对应的共享密钥值 $k_i=e(\Pi(Q(H_1(z_i))),bP)^c$.最后对 z_i 和 k_i 联合求哈希值得到 $K_i^C=H_2(z_i,k_i)$.

7) 计算交集.参与方 C 判断 K_i^C 是否在 K 中,若 $K_i^C\in K$,则输出 z_i.

3.2.2 正确性分析

与 3.1.2 节类似,对于交集中元素 $x_i=y_i=z_i$ 有:

$$H_2(z_i,k_i)=H_2(z_i,e(\Pi(Q(H_1(z_i))),bP)^c)=$$
$$H_2(z_i,e(a_iP,bP)^c)=H_2(z_i,e(P,P)^{a_ibc})=$$
$$H_2(y_i,e(a_iP,cP)^b)=H_2(y_i,$$
$$e(\Pi(Q(H_1(y_i))),cP)^b)=H_2(y_i,k_i).$$

因此,满足 $H_2(y_i,k_i)\in K$ 的元素即是交集 $X\cap Y\cap Z$ 的元素.

3.2.3 安全性证明

考虑到参与方的合谋情况,安全性证明分为 3 种情况,在定理 2~4 中分别讨论.

定理 2. 假设 Π,Π^{-1}是理想置换,协议 2 在恶意参与方 B,C 合谋的情况下是安全的.

证明. 由于诚实参与方并未收到任何来自恶意参与方 B,C 的输入,所以在 B,C 合谋的情况下,不需要考虑敌手对诚实参与方输出的影响,也即不需要进行输入提取.因 Π 是理想置换,每一个 $\Pi^{-1}(a_iP)$ 的输出值都是一个均匀随机的,因此多项式 $Q(\cdot)$与输入 x 无关,$Q(\cdot)$与随机选取的同阶多项式是无法区分的.B,C 未获得任何有用信息,从而证明了在恶意参与方 B,C 合谋的情况下,协议 2 安全地实现了隐私交集计算的功能. 证毕.

定理 3. 假设 H_1, H_2 是随机预言机,Π,Π^{-1}是理想置换,则协议 2 在恶意参与方 A,C 合谋的情况下安全性归约于判定性双线性问题的困难性假设.

证明思路如下:模拟器的主要任务是提取 2 个输入集合 $\overline{X}$ 和 $\overline{Z}$,发送给理想 PSI 函数,获得 $\overline{X}\cap Y\cap\overline{Z}$,然后适当地模拟了消息 K.想要区分合谋参与方作为参与者输出的消息和作为窃听者输出的消息,前者与 $\overline{X}\cap\overline{Z}$ 的元素对应,后者可用随机值代替.诚实的参与方 B 对每个 $y\in Y$ 计算 $\Pi(Q(H_1(y)))$作为一个标准消息,而敌手只有以下情况同时发生时才拥有这个值:它查询了 $H_1(y)$且对随机置换做了反向查询 Π^{-1}得到 $Q(H_1(y))$.若敌手直接选择 $Q(H_1(y))$或只对理想置换做正向查询 Π,则敌手没有对应的值,在模拟器中它将对该结果值没有控制.同理,模拟器观察 H_2 的所有查询,可以识别哪些 K 的实例会给出敌手可以识别的输出,其他输出都可以安全地用随机输出来替换.

证明. 首先刻画模拟器 S 的能力:

1) S 诚实地扮演随机预言机 H_1,H_2 和理想置换 $\Pi^{\pm}$的角色.

① 对于所有敌手发出的查询 $H_1(y)$,若未查询过 y,则 S 随机选取元素 h_1 作为返回,并将(y,h_1)

记录在列表 O_1 中;否则直接返回 O_1 中的对应值.

② 对于每一个敌手发出的查询 $H_2(y,k)$,若未查询过(y,k),则 S 随机选取元素 h_2 作为返回,并将(y,k,h_2)记录在列表 O_2 中;否则直接返回 O_2 中的对应值.

③ 对于每个查询 $\Pi^{-1}(m)$,若未查询过 $\Pi(f)$,则 S 随机选取元素 f 作为返回,并将(f,m)记录在列表 O_Π 中;否则直接返回 O_Π 中的对应值.

2) 在收到多项式 $Q(\cdot)$之后,S 定义集合 $\overline{X}=\{x\mid x\in O_1 \text{ and } Q(H_1(x))\in O_\Pi\}$和 $\overline{Z}=\{z\mid \exists k':(z,e(\Pi(Q(H_1(z))),cP)^b,k')\in O_2\}$,并将 $\overline{X}$ 和 $\overline{Z}$ 发送到理想 PSI 函数.

3) 从理想 PSI 函数接收到 $W=\overline{X}\cap Y\cap\overline{Z}$ 后,S 对所有 $w\in W$ 计算 $k_w=e(\Pi(Q(H_1(w))),cP)^b$,定义 $K=\{H_2(w,k_w)\mid w\in W\}$,然后不断向 K 中添加随机值,直到$|K|=|Y|$.

4) S 最终将 K 发送给敌手.

以下通过混合序列 $hybrid_h$,证明了这个模拟协议与真实协议是无法区分的.

1) $hybrid_0$.这是真实的交互,与参与方 B 按照协议规则诚实地运行,其中 K 定义为

$$K=\{H_2(y,e(\Pi(Q(H_1(y))),cP)^b)\mid y\in Y\},$$

同时,列表 O_1,O_2 和 O_Π 也被生成.

2) $hybrid_1$.与 $hybrid_0$ 唯一的不同是:在计算 k_i 时,若存在 $y\in Y$ 满足 $y\notin O_1$ 但 $Q(H_1(y))\in O_\Pi$,则交互终止.这意味着敌手从未查询 $H_1(y)$,但 $Q(H_1(y))$却是它从 Π^{-1}输出的值.这种概率是可忽略的:对于任意 $f\in O_\Pi$,多项式等式 $Q(\cdot)=f$ 至多有 n 个解且在 F 中均匀分布,因此至少有 $n/|F|$ 概率满足 $Q(H_1(y)=f)$.假设敌手对 H_1 做了共 q 次查询,通过 n 个 $y\in Y$ 和 q 个 $f\in O_\Pi$,总概率为 $n^2q/|F|$,这是可忽略的.

3) $hybrid_{2,i}$.对于每一个 $i\in[q]$,对于形式为 $\Pi(f)=m$ 的前 i 次查询中,若从未查询过 $\Pi^{-1}(m)$,则将 f 添加到集合 S_i.记 S_i 是前 i 次查询中敌手没有的元素对应的 Π 的输出.显然 S_i 和 O_Π 是不相交的,所以计算 K 为

$$K=\{H_2(y,e(\Pi(Q(H_1(y))),cP)^b)\mid y\in Y \text{ and } Q(H_1(y))\notin S_i\}.$$

最后在 K 中加入随机元素,直到$|K|=n$.

此后,$hybrid_{2,i+1}$将 $hybrid_{2,i}$ 中的 $e(\Pi(Q(H_1(y))),cP)^b$ 替换为随机值.根据 DBDH 假设可知,$hybrid_{2,i+1}$和 $hybrid_{2,i}$是无法区分的.

4) $hybrid_3$.重写了 $hybrid_{2,q}$,所有 $\Pi(f)=m$ 都用 S_q 和 O_Π 表示,对于所有满足 $\Pi(f)=m$ 的值必然属于这 2 个集合的其中之一,即 $Q(H_1(y))\notin S_i$ 等价于 $Q(H_1(y))\in O_\Pi$,因此 K 可以表示为

$$K=\{H_2(y,e(\Pi(Q(H_1(y))),cP)^b)\mid y\in Y \text{ and } Q(H_1(y))\in O_\Pi\}.$$

由 $hybrid_1$ 已知,若存在任何 $y\notin O_1$ 但 $Q(H_1(y))\in O_\Pi$,则交互将终止.这意味着 $Q(H_1(y))\in O_\Pi$ 隐含着 $y\in O_1$.因此,K 可以进一步改写为

$$K=\{H_2(y,e(\Pi(Q(H_1(y))),cP)^b)\mid y\in Y\cap O_1 \text{ and } Q(H_1(y))\in O_\Pi\}.$$

5) $hybrid_4$.$hybrid_3$ 基础上,诚实的参与方 B 查询 H_2 以获得 K.若一个 H_2 查询是第 1 次查询,但结果在 O_2 中,则协议终止.此类事件的概率是 $|O_2|/|F|=n/|F|$,是可以忽略的,则 $hybrid_4$ 与 $hybrid_3$ 是无法区分.

假设 $hybrid_4$ 维护了前面描述的列表 O_2,即 $(y,e(\Pi(Q(H_1(y))),cP)^b,k')\in O_2$ 意味着敌手查询 $H_2(y,e(\Pi(Q(H_1(y))),cP)^b)$并得到输出密钥的随机预言机查询结果.由于参与者 B 只识别敌手已经向 H_2 查询过的值,因此 $hybrid_4$ 可以写作:

$$K=\{k'\mid (y,e(\Pi(Q(H_1(y))),cP)^b,k')\in O_2\mid y\in Y\cap O_1 \text{ and } Q(H_1(y))\in O_\Pi\}.$$

假设记 $\overline{X}=\{x\mid x\in O_1 \text{ and } Q(H_1(x))\in O_\Pi\}$及 $\overline{Z}=\{z\mid \exists k':(z,e(\Pi(Q(H_1(z))),cP)^b,k')\in O_2\}$,则 K 等价为

$$K=\{H_2(y,e(\Pi(Q(H_1(y))),cP)^b)\mid y\in \overline{X}\cap Y\cap\overline{Z}\}.$$

此时,$hybrid_4$ 与模拟器的行为是相同的.

综上,模拟协议和真实协议是不可区分的,从而证明了在恶意参与方 A,C 合谋的情况下,协议 2 安全地实现了隐私交集计算功能.　　证毕.

定理 4. 假设 H_1,H_2 是随机预言机,Π,Π^{-1}是理想置换,则协议 2 在恶意参与方 A,B 合谋的情况下安全性归约于判定性双线性问题的困难性假设.

证明思路如下:模拟器 S 的主要任务是提取集合 $\overline{X}$ 和 $\overline{Y}$ 发送给理想函数.在接收到多项式 $Q(\cdot)$之后,诚实的参与方 C 对每个 $z\in Z$ 计算 $\Pi(Q(H_1(z)))$,而敌手只有以下情况发生时才拥有这个值:查询了 $H_1(z)$并且它查询反向随机置换 Π^{-1}得到 $Q(H_1(z))$.若敌手直接选择 $Q(H_1(z))$或直接对理想置换做正向查询 Π,则敌手没有对应的值,S 将

对该结果值没有控制.同理,在接收到敌手给出的 K 后,S 观察所有 H_2 查询,可以知道哪些值通过 H_2 放入了 K 中.同时,Π 由 S 控制,因此 S 可以获得随机置换的输入输出值,进而得知哪些值是 z 对应的正确输出密钥,从而完成输入提取.

证明. 首先刻画模拟器 S 的能力:

1) S 扮演随机预言机 H_1,H_2 和理想置换 $\Pi^{\pm}$ 的角色同定理 3.

2) 在收到多项式 $Q(\cdot)$ 之后,S 定义集合

$$\overline{X}=\{x\mid x\in O_1 \text{ and } Q(H_1(x))\in O_{\Pi}\}.$$

3) 在接收到来自敌手的 K 后,S 定义集合

$$\overline{Y}=\{y\mid \exists k':(y,e(\Pi(Q(H_1(y))),bP)^c,k')\in O_2 \text{ and } k'\in K\}.$$

4) S 将 $\overline{X}$ 和 $\overline{Y}$ 发送到理想 PSI 函数获得 $\overline{X}\cap\overline{Y}\cap Z$.

以下通过混合序列 $hybrid_h$,证明了这个模拟协议与真实协议是无法区分的.

1) $hybrid_0$.这是真实的交互,合谋方 A,B 与参与方 C 按照协议规则运行,交集可以表示为

$$\{z\in Z\mid H_2(z,e(\Pi(Q(H_1(z))),bP)^c)\in K\},$$

同时,列表 O_1,O_2 和 O_{Π} 也被生成.

2) $hybrid_1$.与 $hybrid_0$ 唯一的不同是:$\Pi^{\pm}$ 的模拟.所有对 Π 和 Π^{-1} 的新查询都以随机值进行响应.若 Π 或 Π^{-1} 获得重复的输出,则交互将终止.因此 $hybrid_1$ 与 $hybrid_0$ 是无法区分的.

3) $hybrid_2$.与 $hybrid_1$ 不同是:若存在 $z\in Z$ 满足 $z\notin O_1$ 但 $Q(H_1(z))\in O_{\Pi}$,则交互终止,即敌手从未查询 $H_1(z)$,但 $Q(H_1(z))$ 却是它从 Π^{-1} 输出的值.这种概率是可忽略的.对于任意 $f\in O_{\Pi}$,多项式等式 $Q(\cdot)=f$ 至多有 n 个解且在 F 中均匀分布,因此至少有 $n/|F|$ 概率满足 $Q(H_1(z))=f$.假设敌手对它的预言机做了总共 q 次查询,通过 n 个 $z\in Z$ 和 q 个 $f\in O_{\Pi}$,总概率为 $n^2q/|F|$,这个概率也是可忽略的.若不终止,则意味着 $Q(H_1(z))\in O_{\Pi}$ 且 $z\in O_1$.因此输出形式改写为

$$\{z\in Z\mid H_2(z,e(\Pi(Q(H_1(z))),bP)^c)\in K \text{ and } Q(H_1(z))\in O_{\Pi} \text{ and } z\in O_1\}.$$

4) $hybrid_3$.诚实的接收者查询 H_2 以获得输出,与 $hybrid_2$ 唯一的不同是 H_2 如何模拟.若 H_2 查询是第 1 次被查询但结果却在 K 中,则协议终止.此类事件的概率是 $|K|/|F|=n/|F|$,这是可忽略的,因此 $hybrid_3$ 与 $hybrid_2$ 无法区分.

假设 $hybrid_3$ 维护了列表 O_2,即 $(z,e(\Pi(Q(H_1(z))),bP)^c,k')\in O_2$ 意味着敌手查询 $H_2(z,e(\Pi(Q(H_1(z))),bP)^c)$ 并得到结果 k'.由于 C 识别敌手已经向 H_2 查询过的值,可记作:

$$\{z\in Z\mid \exists k':(z,e(\Pi(Q(H_1(z))),bP)^c,k')\in O_2 \text{ and } k'\in K \text{ and } Q(H_1(z))\in O_{\Pi} \text{ and } z\in O_1\},$$

等价于:

$$Z\cap\{x\mid \exists k':(x,e(\Pi(Q(H_1(x))),bP)^c,k')\in O_2 \text{ and } k'\in K \text{ and } Q(H_1(x))\in O_{\Pi} \text{ and } x\in O_1\}.$$

假设记 $\overline{X}=\{x\mid x\in O_1 \text{ and } Q(H_1(x))\in O_{\Pi}\}$ 和 $\overline{Y}=\{y\mid \exists k':(y,e(\Pi(Q(H_1(y))),bP)^c,k')\in O_2 \text{ and } k'\in K\}$,则交集等价于:

$$Z\cap\{x\mid x\in O_1 \text{ and } Q(H_1(x))\in O_{\Pi}\}\cap\{y\mid \exists k':(y,e(\Pi(Q(H_1(y))),bP)^c,k')\in O_2 \text{ and } k'\in K\}.$$

至此,$hybrid_3$ 输出为 $Z\cap\overline{X}\cap\overline{Y}$,这和理想输出是相同的.

综上,模拟器的行为和真实协议是不可区分的,从而证明了在恶意参与方 A,B 合谋的情况下,协议 2 依然安全地实现了隐私交集计算功能. 证毕.

4 协议性能与比较

4.1 实验环境与参数

本文使用 C++实现了基于密钥协商的三方恶意隐私集合交集计算协议,并与同类方案进行了对比.实验环境为:Ubuntu 18.04.4 LTS, Intel® Core™ i7-8750H CPU @2.20 GHz, 16 GB RAM.实验中采用带有固定密钥且分组长度为 128 b 的 AES 算法作为理想置换,并使用 SHA2 实例化必要的哈希函数.具体实现中采用了 libOTe 库以及 libsodium 库,最后利用 PBC 库提供的 A 类曲线($y^2=x^3+x$)实现了双线性配对.所有计算均采用 PSI 元素长度为 128 b,计算安全参数 $\kappa=128$,统计安全参数 $\lambda=40$.

4.2 性能评估与分析

小集合场景下的性能:本文分别对各参与方集合元素数量为 2^4,2^5,2^6,2^7,2^8,2^9 这 6 种情况对恶意隐私集合交集协议进行了仿真实验测试.

在实施过程中,协议的计算成本包括:参与方 B 和参与方 C 计算 bP 和 cP,以及 n 个经过双线性运算的密钥输出;参与方 A 计算 n 个 a_iP.3 个参与方对 H_1 和 $\Pi^{\pm}$ 各进行 n 次查询;参与方 B 和 C 分别对 H_2 进行 n 次查询;最后参与方 A 必须插值一个多项式,参与方 B,C 在 n 个点上对多项式求值,这都需要 $O(n\,\mathrm{lb}^2 n)$ 个域操作[41].由参与方 C 进行对比

得出交集,各个参与方负载均衡.该协议的总通信成本包括:参与方 B,C 发送的 bP 和 cP;用于描述多项式 $Q(\cdot)$的 n 个域元素;n 个 H_2 输出,每个 2κ 位.

表 1 展示了本文提出的恶意三方 PSI 协议的通信量和各阶段运行时间.本文的协议适合预计算,可将部分计算放在离线阶段进行.在实验测试性能时,本文将协议分为离线阶段和在线阶段:离线阶段主要是参数的生成以及参与方 A 的多项式插值的计算;在线阶段各方交互并得出交集.耗时为在线阶段和离线阶段的运行时间,通信量为所有参与方发送/接收数据的总和.从实验数据可以看出,本文提出的协议十分适用于参与方所具有的集合元素较少的情况.在集合元素较少的情况下,该协议具有较高的运算效率和极低的通信量;在集合元素较多的情况下,由于本文协议涉及双线性配对计算,所以效率有所下降。对比已有三方 PSI 协议,本文协议仍具有最高的通信效率。因此,本文协议适用于网络带宽和通信量均受限的场景.

Table 1　Communication and Running Time of Our Protocol in Small Set Scenarios

表 1　小集合场景下本文方案通信量和运行时间

集合大小	通信量/KB	在线时间/s	离线时间/s
2^4	2.68	0.029	0.044
2^5	5.25	0.041	0.073
2^6	10.37	0.083	0.119
2^7	20.62	0.143	0.246
2^8	41.08	0.306	0.479
2^9	82.12	0.592	0.936

4.3　性能对比

表 2 展示了本文方案与相关工作的综合对比.其中 n 是集合中元素的个数.表 2 中的文献都可以实现三方 PSI 的计算.本文提出的方案分别提供了半诚实和恶意安全性,并且在任意两方合谋时都是安全的.此外,本文在 2 种安全模型下都实现了线性的通性复杂度和计算复杂度并且仅需要两轮通信交互.文献[31]、文献[43]以及文献[42]可提供半诚实的安全性.相比之下,本文协议的半诚实版本具有更低的通信轮数以及通信复杂度和计算复杂度,并允许任意两方合谋.

Table 2　Comparison of Related Semi-Honest and Malicious Three -Party PSI Protocols

表 2　相关半诚实、恶意三方 PSI 协议对比

协议	安全模型	是否合谋	通信轮数	通信复杂度	计算复杂度
文献[42]	半诚实	是	4	$O(n)$	$O(n\,\mathrm{lb}\,n)$
文献[31]-协议 1	半诚实	是	4	$O(n)$	$O(n)$
文献[31]-协议 2	增强半诚实	是	3	$O(n)$	$O(n)$
文献[43]	半诚实	否	8	$O(n\,\mathrm{lb}\,n)$	$O(n)$
本文(PSI-s)	半诚实	是	2	$O(n)$	$O(n)$
文献[42]	恶意	是	7	$O(n\,\mathrm{lb}\,n)$	$O(n^2)$
文献[44]	恶意	是	12	$O(n)$	$O(n\,\mathrm{lb}^2 n)$
文献[32]	恶意	是	8	$O(n\,\mathrm{lb}\,n)$	$O(n)$
文献[45]	恶意	否	5	$O(n\,\mathrm{lb}\,n)$	$O(n)$
文献[35]-协议 1	恶意	否	5	$O(n)$	$O(n)$
文献[35]-协议 2	恶意	是	4	$O(n)$	$O(n)$
本文(PSI-m)	恶意	是	2	$O(n)$	$O(n)$

文献[32,35,42,44-45]都可以在恶意敌手存在的情况下完成工作.文献[45]表明必须有指定的 2 个服务器是不合谋的,所以文献[45]在三方交集中存在两方合谋时是不安全的.文献[42]基于加法同态加密框架构建,且在恶意版本的协议中需要 $O(n^2)$ 计算复杂度,其未提供任何实验数据和代码,但预计其效率要比本文低的多.文献[44]基于不经意线性函数评估(oblivious linear function evaluation, OLE)构建,同样需要大量公钥操作,其优势在于通信量较低,当集合包含 2^{16} 个元素时,有固定通信量为 80 MB,这是本文通信量的 8 倍.文献[32]和文献[35]均基于高效的 OT 扩展,具有较高的运算效率,主要瓶颈均在于通信.

大集合场景下相关方案的通信量对比:文献[32]和文献[35]与本文在 3 个参与方,各参与方集合大小为 2^8,2^{12},2^{16},2^{20},且存在合谋情况下的通信

量对比如表 3 所示.可以看出,拥有相同的集合大小时,相比于文献[32]和文献[35],本文方案的通信量降低了 89%～98%.因此,本文的协议也将适用于此类具有较大集合(2^{10}～2^{20}个元素)且需要降低通信量的场景中.例如在集合大小是 2^{20} 时,文献[32]和文献[35]通信开销分别约为 20 GB 和 1.6 GB,这在通信带宽受限的场景中是不可接受的.

Table 3 Communication Comparison of Malicious Secure Three-Party PSI Protocol in Large Set Scenarios

表 3 大集合场景下恶意安全的三方 PSI 协议通信量对比

MB

协议	集合大小			
	2^8	2^{12}	2^{16}	2^{20}
文献[32]	14.45	214.15	1 623.28	20 876.96
文献[35]	0.34	5.52	95.07	1 669.17
本文(PSI-m)	0.04	0.64	10.25	164.01

小集合场景下相关方案的对比:基于同态加密的方案同样适应于弱通信场景,具有较低通信量,但由于需要大量公钥加密操作,计算速度比本文方案慢几个数量级.文献[25]与文献[26]均基于同态加密实现了多方 PSI,这 2 个文献与本文方案的通信量和运行时间对比如图 4 所示.由图 4 可知,文献[26]的通信量和运行时间都很高,这与其采用同态加密的布隆过滤器有关;文献[25]通信量较低,与本文基本相同,但其总运行时间是本文的 10～25 倍.

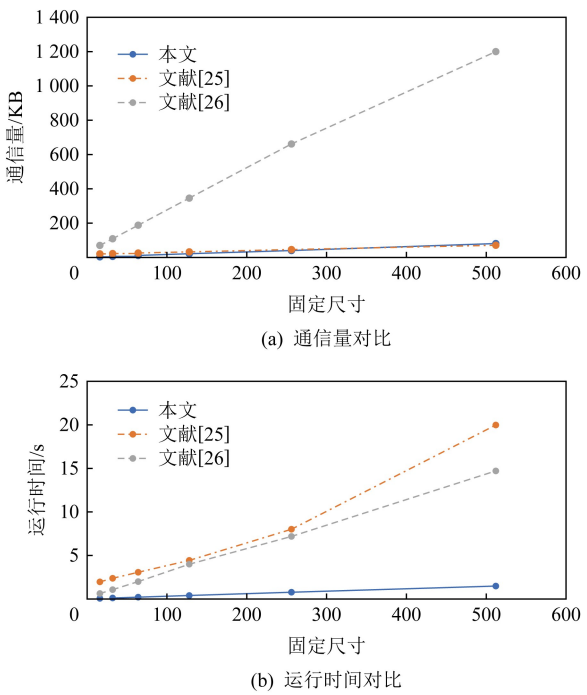

Fig. 4 Comparison with HE-based PSI in small setscenarios

图 4 小集合场景下与基于同态加密的 PSI 方案对比

5 结 论

PSI 协议是安全多方计算的重点研究问题之一.本文提出了 2 个基于密钥协商的三方隐私集合交集计算协议,分别可以抵抗半诚实和恶意敌手且允许任意两方合谋,通过模拟范式证明了方案的安全性.与现有方案相比,本文的方案具有较低的通信代价和更高的安全性,尤其适用于带有小集合的三方隐私集合交集计算场景.在未来的工作中,我们将考虑进一步提高计算效率并扩展到普适多方的场景,一种可能的优化是通过寻找不同的编码方式以降低多项式插值过程的计算成本.

作者贡献声明:张蕾负责论文思路构建和框架设计;贺崇德负责撰写论文和实验;魏立斐提出指导意见并负责论文校对与修订.

参 考 文 献

[1] Shen Liyan, Chen Xiaojun, Shi Jinqiao, et al. Survey on private preserving set intersection technology [J]. Journal of Computer Research and Development, 2017, 54(10): 2153-2169 (in Chinese)
(申立艳, 陈小军, 时金桥, 等. 隐私保护集合交集计算技术研究综述[J]. 计算机研究与发展, 2017, 54(10): 2153-2169)

[2] Wei Lifei, Liu Jihai, Zhang Lei, et al. A survey of privacy preserving oriented set intersection computation [J]. Journal of Computer Research and Development, 2022, 59(8): 1782-1799 (in Chinese)
(魏立斐, 刘纪海, 张蕾, 等. 面向隐私保护的集合交集计算综述[J]. 计算机研究与发展, 2022, 59(8): 1782-1799)

[3] Yung M. From mental poker to core business: Why and how to deploy secure computation protocols? [C] //Proc of the 22nd ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2015: 1-2

[4] Privacy-preserving Data Mining: Models and Algorithms [M]. Berlin: Springer, 2008

[5] Demmler D, Rindal P, Rosulek M, et al. PIR-PSI: Scaling private contact discovery [J]. Proceedings on Privacy Enhancing Technologies, 2018, 2018(4): 159-178

[6] Duong T, Phan D H, Trieu N. Catalic: Delegated PSI cardinality with applications to contact tracing [C] //Proc of the 26th Int Conf on the Theory and Application of Cryptology and Information Security. Berlin: Springer, 2020: 870-899

[7] Ion M, Kreuter B, Nergiz E, et al. Private intersection-sum protocol with applications to attributing aggregate ad conversions [J/OL]. Cryptology ePrint Archive, 2017 [2020-11-30]. https://eprint.iacr.org/2017/738

[8] Rabin M O. How to exchange secrets with oblivious transfer [J/OL]. Cryptology ePrint Archive, 2005 [2021-12-30]. https://eprint.iacr.org/2005/187

[9] Kolesnikov V, Kumaresan R, Rosulek M, et al. Efficient batched oblivious PRF with applications to private set intersection [C] //Proc of the 23rd ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2016: 818-829

[10] Pinkas B, Rosulek M, Trieu N, et al. Spot-light: Lightweight private set intersection from sparse ot extension [C] //Proc of the 39th Annual Int Cryptology Conf. Berlin: Springer, 2019: 401-431

[11] Chase M, Miao Peihan. Private set intersection in the Internet setting from lightweight oblivious PRF [C] //Proc of the 40th Annual Int Cryptology Conf. Berlin: Springer, 2020: 34-63

[12] Pinkas B, Schneider T, Zohner M. Faster private set intersection based on OT extension [C] //Proc of the 23rd USENIX Security Symp. Berkeley. CA: USENIX Association, 2014: 797-812

[13] Pinkas B, Schneider T, Segev G, et al. Phasing: Private set intersection using permutation-based hashing [C] //Proc of the 24th USENIX Security Symp. Berkeley, CA: USENIX Association, 2015: 515-530

[14] Rindal P, Rosulek M. Malicious-secure private set intersection via dual execution [C] //Proc of the 24th ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2017: 1229-1242

[15] Wei Lifei, Wang Qin, Zhang Lei, et al. Efficient private set intersection protocols with semi-trusted cloud server aided [J/OL]. Journal of Software, 1-13 [2022-06-15]. http://www.jos.org.cn/1000-9825/6397.html (in Chinese)
(魏立斐, 王勤, 张蕾, 等. 半可信云服务器辅助的高效隐私交集计算协议[J/OL]. 软件学报, 1-13 [2022-06-15]. http://www.jos.org.cn/1000-9825/6397.html)

[16] Ion M, Kreuter B, Nergiz E, et al. Private intersection-sum protocol with applications to attributing aggregate ad conversions [J/OL]. Cryptology ePrint Archive, 2017[2020-11-30]. https://eprint.iacr.org/2017/738

[17] Huberman B, Franklin M, Hogg T. Enhancing privacy and trust in electronic communities [C] //Proc of the 1st ACM Conf on Electronic Commerce. New York: ACM, 1999: 78-86

[18] Cristofaro E D, Kim J, Tsudik G. Linear-complexity private set intersection protocols secure in malicious model [C] //Proc of the 16th Int Conf on the Theory and Application of Cryptology and Information Security. Berlin: Springer, 2010: 213-231

[19] Jarecki S, Liu Xiaomin. Fast secure computation of set intersection [C] //Proc of the 7th Int Conf on Security and Cryptography for Networks. Berlin: Springer, 2010: 418-435

[20] Rosulek M, Trieu N. Compact and malicious private set intersection for small sets [C] //Proc of the 27th ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2021: 1166-1181

[21] Mansy D, Rindal P. Endemic oblivious transfer [C] //Proc of the 26th ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2019: 309-326

[22] Cristofaro E D, Tsudik G. Practical private set intersection protocols with linear complexity [C] //Proc of the 14th Int Conf on Financial Cryptography and Data Security. Berlin: Springer, 2010: 143-159

[23] Cong K, Moreno R C, da Gama M B, et al. Labeled PSI fromhomomorphic encryption with reduced computation and communication [C] //Proc of the 27th ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2021: 1135-1150

[24] Chen Hao, Huang Zhicong, Laine K, et al. Labeled PSI from fully homomorphic encryption with malicious security [C] //Proc of the 25th ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2018: 1223-1237

[25] Bay A, Erkin Z, Alishahi M, et al. Multi-party private set intersection protocols for practical applications [C] //Proc of the 18th Int Conf on Security and Cryptography. Setubal: SciTePress, 2021: 515-522

[26] Bay A, Erkin Z, Hoepman J H, et al. Practical multi-party private set intersection protocols [J]. IEEE Transactions on Information Forensics and Security, 2021, 17: 1-15

[27] Heinrich A, Hollick M, Schneider T, et al. PrivateDrop: Practical privacy-preserving authentication for Apple AirDrop [C] //Proc of the 30th USENIX Security Symp. Berkeley, CA: USENIX Association, 2021: 3577-3594

[28] Pinkas B, Rosulek M, Trieu N, et al. PSI from PaXoS: Fast, malicious private set intersection [C] //Proc of the 39th Annual Int Conf on the Theory and Applications of Cryptographic Techniques. Berlin: Springer, 2020: 739-767

[29] Rindal P, Schoppmann P. VOLE-PSI: Fast OPRF and circuit-PSI from vector-OLE [C] //Proc of the 40th Annual Int Conf on the Theory and Applications of Cryptographic Techniques. Berlin: Springer, 2021: 901-930

[30] Boyle E, Couteau G, Gilboa N, et al. Efficient two-round OT extension and silent non-interactive secure computation [C] //Proc of the 26th ACM SIGSAC Conf on Computer and Communications Security.New York: ACM, 2019: 291-308

[31] Kolesnikov V, Matania N, Pinkas B, et al. Practical multi-party private set intersection from symmetric-key techniques [C] //Proc of the 24th ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2017: 1257-1272

[32] Efraim A B, Nissenbaum O, Omri E, et al. PSImple: Practical multiparty maliciously-secure private set intersection [R/OL]. Cryptology ePrint Archive, 2021 [2021-05-13]. https://eprint.iacr.org/2021/122

[33] Inbar R, Omri E, Pinkas B. Efficient scalable multiparty private set-intersection via garbled Bloom filters [C] //Proc of the 15th Int Conf on Security and Cryptography for Networks. Berlin: Springer, 2018: 235-252

[34] Rindal P, Rosulek M. Improved private set intersection against malicious adversaries [C] //Proc of the 36th Annual Int Conf on the Theory and Applications of Cryptographic Techniques. Berlin: Springer, 2017: 235-259

[35] Nevo O, Trieu N, Yanai A. Simple, fast malicious multiparty private set intersection [C] //Proc of the 27th ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2021: 1151-1165

[36] Garimella G, Pinkas B, Rosulek M, et al. Oblivious key-value stores and amplification for private set intersection [C] //Proc of the 41st Annual Int Cryptology Conf. Berlin: Springer, 2021: 395-425

[37] Bellare M, Hoang V T, Keelveedhi S, et al. Efficient garbling from a fixed-key blockcipher [C] //Proc of the 34th IEEE Symp on Security and Privacy. Piscataway, NJ: IEEE, 2013: 478-492

[38] Guo Chun, Katz J, Wang Xiao, et al. Efficient and secure multiparty computation from fixed-key block ciphers [C] // Proc of the 41st IEEE Symp on Security and Privacy. Piscataway, NJ: IEEE, 2020: 825-841

[39] Boneh D, Franklin M. Identity-based encryption from the Weil pairing [C] //Proc of the 21st Annual Int Cryptology Conf. Berlin: Springer, 2001: 213-229

[40] Joux A. A one round protocol for tripartite Diffie-Hellman [C] //Proc of the 4th Int Algorithmic Number Theory Symp. Berlin: Springer, 2000: 385-393

[41] Moenck R, Borodin A. Fast modular transforms via division [C] //Proc of the 13th Annual Symp on Switching and Automata Theory. Piscataway, NJ: IEEE, 1972: 90-96

[42] Hazay C, Venkitasubramaniam M. Scalable multi-party private set-intersection [C] //Proc of the 20th IACR Int Conf on Practice and Theory in Public Key Cryptography. Berlin: Springer, 2017: 175-203

[43] Chandran N, Dasgupta N, Gupta D, et al. Efficient linear multiparty PSI and extensions to circuit/quorum PSI [C] // Proc of the 27th ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2021: 1182-1204

[44] Ghosh S, Nilges T. An algebraic approach to maliciously secure private set intersection [C] //Proc of the 38th Annual Int Conf on the Theory and Applications of Cryptographic Techniques. Berlin: Springer, 2019: 154-185

[45] Zhang En, Liu Fenghao, Lai Qiqi, et al. Efficient multi-party private set intersection against malicious adversaries [C] // Proc of the 27th ACM SIGSAC Conf on Cloud Computing Security Workshop. New York: ACM, 2019: 93-104

Zhang Lei, born in 1983. PhD, associate professor. Member of CCF. Her main research interests include cryptography, data security and access control.

张 蕾, 1983 年生.博士,副教授.CCF 会员.主要研究方向为密码学、数据安全、访问控制.

He Chongde, born in 1997. Master candidate. Student member of CCF His main research interests include cryptography and information security.

贺崇德, 1997 年生.硕士研究生.CCF 学生会员.主要研究方向为密码学、信息安全.

Wei Lifei, born in 1982. PhD, professor, master supervisor. Senior member of CCF. His main research interests include information security, privacy preserving and cryptography.

魏立斐, 1982 年生.博士,教授,硕士生导师.CCF 高级会员.主要研究方向为信息安全、隐私保护、密码学.