Computer Science and Application 计算机科学与应用, 2025, 15(4), 80-86
Published Online April 2025 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2025.154080

Hans汉斯

云辅助阈值多方隐私集合交集

刘芗宇，岳笑含

沈阳工业大学信息科学与工程学院，辽宁 沈阳

收稿日期：2025年3月1日；录用日期：2025年3月31日；发布日期：2025年4月8日

摘　要

隐私集合交集(Private Set Intersection, PSI)协议是一种具有重要实际意义的安全多方计算协议，广泛应用于多方私有输入集合求交集的场景。阈值多方PSI协议作为PSI协议的一种灵活形式，能够适应更多复杂场景。本文给出了一种一次的云辅助阈值多方PSI模型(Cloud-assisted Threshold Multi-party Private Set Intersection, CTMPSI)，旨在优化发送方在资源受限场景下的性能。该协议通过引入云服务器辅助计算，显著降低了发送方的计算和通信开销，同时在半诚实模型下确保了输入集合元素的隐私性。此外，CTMPSI实现了发送方上传加密数据后即可离线的功能，进一步提升了协议的实用性。本文详细描述了CTMPSI协议的设计框架和性能评估。实验结果表明，在不平衡输入集合场景中，CTMPSI协议相较于现有的多方PSI协议，在性能上取得了显著提升。该协议为资源受限场景下的阈值多方PSI应用提供了高效且安全的解决方案，具有重要的理论价值和实际意义。

关键词

隐私集合交集，云辅助，资源受限，同态加密，秘密分享

Cloud-Assisted Threshold Multi-Party Private Set Intersection

Xiangyu Liu, Xiaohan Yue

School of Information Science and Engineering, Shenyang University of Technology, Shenyang Liaoning

Received: Mar. 1st, 2025; accepted: Mar. 31st, 2025; published: Apr. 8th, 2025

Abstract

Private Set Intersection (PSI) protocol is a secure multi-party computation protocol with significant practical applications, widely used in scenarios where multiple parties need to compute the intersection of their private input sets. As a flexible variant of PSI, threshold multi-party PSI can adapt to more complex scenarios. This paper proposes a one-round cloud-assisted threshold multi-party PSI model (Cloud-assisted Threshold Multi-party Private Set Intersection, CTMPSI), aiming to optimize

文章引用: 刘芗宇, 岳笑含. 云辅助阈值多方隐私集合交集[J]. 计算机科学与应用, 2025, 15(4): 80-86.
DOI: 10.12677/csa.2025.154080

the performance of senders in resource-constrained scenarios. By introducing cloud server-assisted computation, the protocol significantly reduces the computational and communication overhead for senders while ensuring the privacy of input set elements in the semi-honest model. Additionally, CTMPSI enables senders to go offline after uploading encrypted data, further enhancing the practicality of the protocol. This paper provides a detailed description of the design framework and performance evaluation of CTMPSI. Experimental results demonstrate that, in scenarios with unbalanced input sets, CTMPSI achieves significant performance improvements compared to existing multi-party PSI protocols. The protocol offers an efficient and secure solution for threshold multi-party PSI applications in resource-constrained environments, holding important theoretical and practical significance.

Keywords

Private Set Intersection, Cloud-Assisted, Resource-Constrained, Homomorphic Encryption, Secret Sharing

1. 引言

随着信息技术的飞速发展，网络技术的革新深刻改变了人类社会的运作模式。这种技术演进在为日常生活带来前所未有的便利的同时，也催生了一系列新型社会问题。在数据驱动的时代背景下，个体日常行为在网络上产生的信息量呈现指数级增长态势，而数据分析技术使得个人敏感信息的提取成为可能，导致隐私泄露风险显著攀升。为应对这一挑战，我国相继颁布了多项数据保护相关法规[1]，如《中华人民共和国个人信息保护法》《中华人民共和国网络安全法》以及《中华人民共和国数据安全法》等，致力于构建完善的数据治理体系。在此背景下，如何在确保数据安全的前提下实现数据的有效利用，已成为当前亟待解决的关键问题。

隐私集合交集(private set intersection, PSI)协议是一种安全多方计算协议，旨在保护参与方的输入集合中元素隐私的同时输出其交集[2]。作为密码学工具之一，PSI 协议在多个领域中广泛应用。例如，在私有通讯录发现[3]、隐私保护数据挖掘[4]、衡量在线广告转化率[5]和接触追踪[6]等场景中也发挥着重要作用。

传统两方 PSI 协议在处理多方参与的场景时存在明显局限性。为应对这一挑战，多方私有集合交集(Multi-party PSI, MPSI)协议[7]应运而生。MPSI 协议能够在多个参与方不泄露各自隐私输入集合的前提下，计算多方输入集合的交集。为进一步提升协议的灵活性，研究者提出了多方门限 PSI (Threshold MPSI, T-MPSI)协议[8]。T-MPSI 协议支持在 n 名参与方中识别出由 $t\left(t\le n\right)$ 名参与者共同持有的数据，这一特性使其在群众满意度调查、号码标记等现实场景中具有广泛应用潜力。传统 MPSI 协议或 T-MPSI 协议在处理此类问题时，仅增加了资源受限发送方的负担，还可能导致发送方需要长时间等待交互结果。因此，降低发送方的通信次数、减少通信和计算复杂度已成为 T-MPSI 协议研究的关键挑战。

2. 技术背景

2.1. 秘密共享

秘密共享在密码学中有着重要的作用，其在安全多方计算协议有着重要的应用。$\left(t,\ n\right)$ 秘密共享是：

一个秘密持有者将一个秘密 S 分成 n 份，即 $s_i\left(i\in[n]\right)$ 并通过安全信道分别发送给参与方 $P_i\left(i\in[n]\right)$，其中任意 t 名参与方合作即可恢复秘密 S。

秘密持有者随机生成 $(t-1)$ 随机数 $r_j\left(j\in[t-1]\right)$，并利用其持有的秘密 S 和随机数 r_j 生成一个 $(t\times1)$ 的列向量 $\vec{v}$，$\vec{v}=\left(S,\ r_i,\ r_2,\cdots,\ r_{t-1}\right)^{\mathrm{T}}$。

秘密持有者生成一个 $(n\times t)$ 的矩阵 M 用于分发秘密，矩阵 M 每一行 $i\left(i\in[n]\right)$ 对应着一名参与者 P_i，其对应着一个 $(t\times1)$ 的行向量 $\left(i^0,\ i^1,\cdots,\ i^{t-1}\right)$，$n$ 名参与者对应的行向量共同构成了矩阵 M，如公式(1)所示：

$$M=\begin{bmatrix}1 & 1 & 1 & \cdots & 1\\ 1 & 2 & 2^2 & \cdots & 2^{t-1}\\ 1 & 3 & 3^2 & \cdots & 3^{t-1}\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & n & n^2 & \cdots & n^{t-1}\end{bmatrix}_{n\times t} \tag{1}$$

秘密持有者分别计算 $M_{i\times t}\cdot\vec{v}=s_i$，$s_i$ 作为秘密分片，分别发送给参与方 P_i。

随机选择 t 参与方，利用其对应 M 中的行向量和其手中的秘密分片，共同恢复秘密值 S，其过程如下：

根据 t 名参与方对应 M 中的行向量构成新的矩阵 $\mathrm{M}_{t\times t}$，求出该矩阵的逆矩阵 $\mathrm{M}_{t\times t}^{-1}$；根据 t 名参与方手中的秘密分片 $s_i\left(i\in[n]\right)$；构成一个 $(t\times1)$ 的列向量 $\vec{s}$，$\vec{s}=\left(s_{1'},\ s_{2'},\cdots,\ s_{t'}\right)^{\mathrm{T}}$；计算 $\left[1,\ 0,\ 0,\cdots,\ 0\right]_{1\times t}\cdot M_{t\times t}^{-1}\cdot\vec{s}=\left[1,\ 0,\ 0,\cdots,\ 0\right]_{1\times t}\cdot\vec{v}=S$。

通过上述秘密分发与重构过程，本文成功实现了秘密的分片与恢复。该过程确保了最多 $(t-1)$ 名参与方合谋，也无法通过其持有的秘密分片和矩阵 M 的相关信息恢复秘密值 S，确保了秘密共享方案在面对部分参与方合谋时的安全性。

2.2. 双线性配对

双线性配对(Bilinear Pairing)，也称双线性映射，是现代密码学方案设计中的一项重要工具，广泛应用于隐私保护计算等领域。具体而言，设 $\mathbb{G}_1,\ \mathbb{G}_2,\ \mathbb{G}_T$ 是三个大素数 q 阶的乘法循环群，其中 $g_1\in\mathbb{G}_1$ 和 $g_2\in\mathbb{G}_2$ 分别为 $\mathbb{G}_1$ 和 $\mathbb{G}_2$ 的生成元。双线性配对的定义为一个映射 $e:\mathbb{G}_1\times\mathbb{G}_2\to\mathbb{G}_T$，且满足下列性质：

1) 双线性：对于任意随机选取的 $a,\ b\in\mathbb{Z}_q$ 和任意的 $g_1\in\mathbb{G}_1,\ g_2\in\mathbb{G}_2$，则有 $e\left(g_1^a,\ g_2^b\right)=e\left(g_1,\ g_2\right)^{a\cdot b}$，这一性质使得双线性配对能够支持复杂的密码学操作。

2) 非退化性：存在 $g_1\in\mathbb{G}_1,\ g_2\in\mathbb{G}_2$，使得 $e\left(g_1,\ g_2\right)\neq1$，其中 1 是 $\mathbb{G}_T$ 中的单位元。

3) 可计算性：对于任意的 $g_1\in\mathbb{G}_1,\ g_2\in\mathbb{G}_2$，高效的多项式时间算法可以计算 $e\left(g_1,\ g_2\right)$，这一性质使得双线性配对在实际应用中具有较高的计算效率。

3. 方案构建

3.1. 方案系统模型

本章提出的云辅助阈值多方 PSI 协议的系统模型如图 1 所示，其中包含 5 个阶段，第一个阶段为接收方和云服务器将公钥发送给发送方，第二个阶段中发送方将其密态输入集合发送给接收方，第三个阶段接收方将发送方的密态集合授权给云服务器用于云服务器计算交集，第四个阶段云服务器将交集的密文发送给接收方，第五个阶段接收方对密文交集密文进行解密并验证，如果验证通过，则输出该密文对应的元素信息。其中协议的参与方包含发送方 $\mathcal{S}$，接收方 $\mathcal{R}$ 和云服务器 $\mathcal{C}$。

发送方：发送方作为资源受限的设备，持有一个规模较小的输入集合 $\mathbb{X}=\{x_1, x_2, \cdots, x_n\}$，其中集合大小为 n。发送方根据接收到的公钥，对其本地集合中的元素进行加密，并将加密后的数据发送至接收方处，以便后续接收方授权给云服务器能够基于密文执行集合交集计算。

接收方：接收方持有一个私钥，并利用秘密分享的方式将该私钥对应的秘密分片分别发送给每个参与方，将使用的矩阵 M 放到广播版上。接收方对云服务器计算交集后发送的密文进行解密，最终获得 t 名参与方都具有的输入集合的元素。

云服务器：云服务器负责接收来自发送方的加密数据，并基于这些密文执行集合交集计算。在计算完成后，获得的交集大小达到 t 后且该 t 名参与方不重复后发送的密文，云服务器利用秘密分片的矩阵 M 对密文计算并聚合，将最后密文发送给接收方。

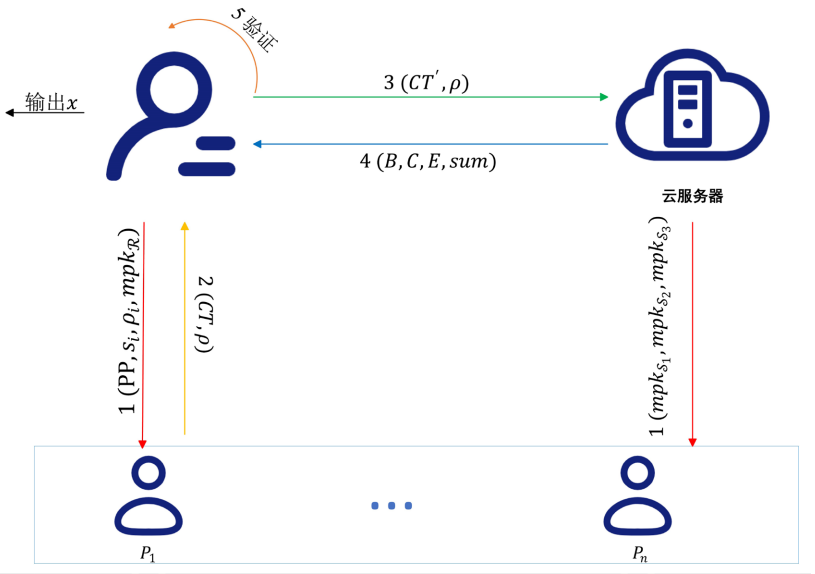

Figure 1. CTMPSI system model
图 1. CTMPSI 系统模型

3.2. 方案形式化定义

1) 初始化阶段

该阶段旨在生成公共参数、秘密份额、接收方 $\mathcal{R}$ 的私钥和公钥、云服务器 $\mathcal{C}$ 的私钥和公钥，并将接收方 $\mathcal{R}$ 和云服务器 $\mathcal{C}$ 的公钥发送给 $\mathcal{S}_i$。在该阶段主要由接收方 $\mathcal{R}$ 执行 PPSetup 算法和 Share 算法，云服务器 $\mathcal{C}$ 执行 CSetup 算法。Share 算法分发秘密份额的分发和接收方的公钥，$\mathrm{CSetup}(\mathrm{PP})$ 算法分发云服务器 $\mathcal{C}$ 的公钥，算法具体内容如下：

$\mathrm{PPSetup}(1^{\lambda}) \to \mathrm{PP}$：输入一个安全参数 λ ,输出公共参数。

$\mathrm{Share}(n, t, \mathrm{PP}) \to \left(sk_{\mathcal{S}}, pk_{\mathcal{R}}, (s_1, \rho(1)), (s_2, \rho(2)), \cdots, (s_n, \rho(n))\right)$：输入发送方的个数 n，阈值 t，公共参数 PP，其中 $n \geq t$，输出接收方 $\mathcal{R}$ 的公私钥和发送方的秘密分片。

$\mathrm{CSetup}(\mathrm{PP}) \to (sk_{\mathcal{C}}, pk_{\mathcal{C}})$：输入公共参数 PP，输出云服务器 $\mathcal{C}$ 的公私钥。

2) 上传阶段

该阶段旨在发送方 $\mathcal{S}_i\ (i \in [n])$ 上传密文，由每个发送方 $\mathcal{S}_i$ 执行 Encrypt 算法。

$\mathrm{Encrypt}(pk_{\mathcal{R}}, pk_{\mathcal{C}}, \mathbb{X}, \mathrm{PP}) \to (CT_{i,j})$：该协议输入接收方 $\mathcal{R}$ 的公钥 $pk_{\mathcal{R}}$，服务器 $\mathcal{C}$ 的公钥 $pk\_\mathcal{C}$，发

送方 $\mathcal{S}_i$ 的输入集合 $\mathbb{X}_i=\left\{x_{i,1}, x_{i,2}, \cdots, x_{i,m_i}\right\}$ 和公共参数 PP。其中 $\left|\mathbb{X}_i\right|=m_i$，$m_i$ 表示发送方 $\mathcal{S}_i$ 的输入集合大小。对于发送方 $\mathcal{S}_i$ 输入集合中的元素 $x_{i,j}\left(j\in\left[m_i\right]\right)$，输出发送方的密态输入集合 $CT_{i,j}$。

3) 授权阶段

该阶段旨在接收方 $\mathcal{R}$ 对接收到的密文进行计算，使得服务器 $\mathcal{C}$ 能够进行交集计算。由接收方 $\mathcal{R}$ 执行 RerandPK 算法。

$\text{RerandPK}\left(CT_i, \text{PP}\right)\rightarrow\left(CT_i^{'}\right)$：输入从发送方 i 接收的密文 $CT_{i,j}$。随机选择私钥 $sk_{\mathcal{R}}\in\mathbb{Z}_p$，对于密文 $CT_{i,j}$ 进行计算从而构建新的 $CT_i^{'}$，将来自每个参与方 i 的全部密文 $CT_i^{'}$ 发送给服务器 $\mathcal{C}$，记作 CT，完成对于服务器 $\mathcal{C}$ 计算交集的授权。

4) 交集阶段

该阶段旨在判断那些元素出现次数达到阈值，主要由服务器 $\mathcal{C}$ 执行 Compare 算法。

$\text{Compare}\left(CT_i^{'}, t, \text{PP}\right)$：输入密文集合 $CT_i^{'}$、阈值 t、公共参数 PP，输出大小为 t 的交集集合 $\mathcal{C}$。

5) 重构阶段

该阶段旨在生成部分解密密钥，用于后续解密满足阈值的密文。该阶段由服务器 $\mathcal{C}$ 执行 Recon 算法。

$\text{Recon}\left(C, \text{M}\right)\rightarrow CT$：输入用于秘密共享的矩阵 M 和大小为 t 的交集集合 $\mathbb{C}$，生成用于阈值解密的密文 CT。

6) 解密阶段

该阶段为满足阈值的密文进行解密。主要由接收方 $\mathcal{R}$ 执行 Decrypt 算法。

$\text{Decrypt}\left(CT, sk_{\mathcal{R}}\right)\rightarrow x$：输入重构阶段的密文 CT 和云服务器的私钥 sk_R，对密文集合进行解密，输出交集集合信息 x。

3.3. 设计目标

(1) 发送方友好性：发送方仅需发送一条消息，且其计算代价和通信成本与接收方的输入集合大小无关。

(2) 离线支持：发送方在上传其输入集合后即可离线，减少其参与成本。

(3) 隐私保护：该协议需确保发送方输入集合的隐私性得到充分保护。具体来说，仅接收方能够获取双方输入集合的交集结果，而云服务器在整个计算过程中无法获取任何有效信息。此外，即使云服务器通过外部渠道获取了发送方部分密文对应的元素信息，也无法对其余发送方输入集合不等于该元素的其他元素隐私构成威胁，从而防止潜在的隐私泄露风险。

(4) 高效性：通过协议的整体优化，显著降低接收方和发送方的通信成本和计算开销。该协议发送方的计算和通信成本与较大的接收方输入集合无关，同时接收方将复杂的计算委托给云服务器的计算资源以减少计算时间，并在该部分云服务器无需接收方的密态输入集合，减少了通信成本。

4. 性能分析

实验采用 RELIC 密码学库，在搭载 2.2 GHz Intel(R) Core(TM) i7-8750H 处理器和 4GB RAM 的 Linux 平台上，通过运行 C 语言实现的实验代码进行协议性能研究。实验设置安全参数为 $\lambda=128$，并基于 BLS12-381 配对友好曲线实现本章协议，以确保实验环境的安全性和可靠性。

本实验将 CTMPSI 协议与适用于非对称输入集合的多方 PSI 协议[9]在发送方的计算时间方面进行对比分析。实验中，接收方的输入集合大小从 2^{10} 到 2^{13} 逐步递增，而发送方的输入集合大小则从 2^4 到 2^9 变化，图 2 展示了 CTMPSI 协议与文献[9]协议在发送方总计算时间上的对比结果，进一步验证了 CTMPSI

协议在非对称输入集合场景下的性能优势。

如图 2 所示，实验通过两个渐变色曲面分别展示了文献[9]中发送方的总计算时间与 CTMPSI 协议中发送方的计算时间，其中曲面颜色随数值的增大逐渐加深，以直观反映计算时间的变化趋势。从图中可以看出，文献[9]中发送方的计算时间始终高于 CTMPSI 协议的计算时间。当发送方和接收方的输入集合规模较小时，两者之间的时间差距相对较小；然而，随着接收方输入集合规模的增加，这一差距逐渐扩大。特别是在发送方输入集合规模显著增大时，两者之间的计算时间差异进一步加剧。尽管如此，这两种多方协议在以上条件下均能够在 30 秒内完成发送方的计算任务，体现了其在实践中的高效性与实用性，具备良好的实际应用价值。

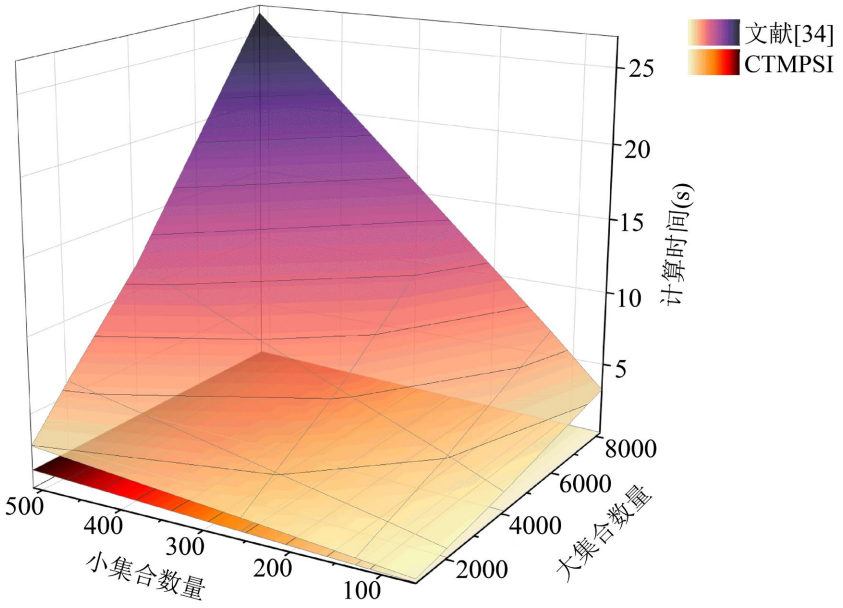

Figure 2. Sender computation time
图 2. 发送方计算时间

参考文献

[1] 黄翠婷, 张帆, 孙小超, 等. 隐私集合求交技术的理论与金融实践综述[J]. 信息通信技术与政策, 2021, 47(6): 50-56.

[2] 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] Hetz, L., Schneider, T. and Weinert, C. (2024) Scaling Mobile Private Contact Discovery to Billions of Users. In: Tsudik, G., Conti, M., Liang, K. and Smaragdakis, G., Eds., *Computer Security—ESORICS* 2023, Springer, 455-476. https://doi.org/10.1007/978-3-031-50594-2_23

[4] Ruan, O., Huang, X. and Mao, H. (2020) An Efficient Private Set Intersection Protocol for the Cloud Computing Environments. 2020 *IEEE 6th Intl Conference on Big Data Security on Cloud* (*BigDataSecurity*), *IEEE Intl Conference on High Performance and Smart Computing*, (*HPSC*) *and IEEE Intl Conference on Intelligent Data and Security* (*IDS*), Baltimore, 25-27 May 2020, 254-259. https://doi.org/10.1109/bigdatasecurity-hpsc-ids49724.2020.00053

[5] Ion, M., Kreuter, B., Nergiz, E., *et al*. (2017) Private Intersection-Sum Protocol with Applications to Attributing Aggregate Ad Conversions. Cryptology ePrint Archive. https://eprint.iacr.org/2017/738

[6] Yang, X., Zhao, Y., Zhou, S. and Wang, L. (2024) A Lightweight Delegated Private Set Intersection Cardinality Protocol. *Computer Standards & Interfaces*, **87**, Article ID: 103760. https://doi.org/10.1016/j.csi.2023.103760

[7] Inbar, R., Omri, E. and Pinkas, B. (2018) Efficient Scalable Multiparty Private Set-Intersection via Garbled Bloom Filters. In: Catalano, D. and De Prisco, R., Eds., *Security and Cryptography for Networks*, Springer, 235-252. https://doi.org/10.1007/978-3-319-98113-0_13

[8] Zhou, J., Su, D. and Deng, J. (2023) Multi-Party Threshold Private Set Intersection Cardinality Based on Encrypted Bloom Filter. 2023 *IEEE International Conferences on Internet of Things* (*iThings*) *and IEEE Green Computing & Communications* (*GreenCom*) *and IEEE Cyber*, *Physical & Social Computing* (*CPSCom*) *and IEEE Smart Data* (*SmartData*) *and IEEE Congress on Cybermatics* (*Cybermatics*), Danzhou, 17-21 December 2023, 503-511. https://doi.org/10.1109/ithings-greencom-cpscom-smartdata-cybermatics60724.2023.00098

[9] Bay, A., Erkin, Z., Hoepman, J., Samardjiska, S. and Vos, J. (2022) Practical Multi-Party Private Set Intersection Protocols. *IEEE Transactions on Information Forensics and Security*, **17**, 1-15. https://doi.org/10.1109/tifs.2021.3118879