隐私集合运算中的关键技术研究

毕业论文预答辩

学　　生：张响鸽

指导老师：陈　宇

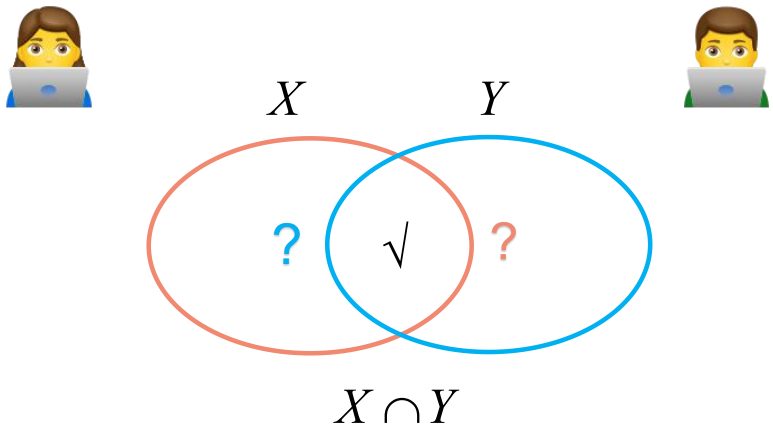
山东大学 网络空间安全学院（研究院）

2023 年 3 月 10 日

PSO 背景

隐私集合运算（Private Set Operation，PSO）是安全多方计算在集合运算场景下的专用协议，可以在保证各个参与方输入隐私的情况下实现各类集合操作，包括交集、并集，以及交集计算等。

- 以隐私集合求交集为例：



- 应用场景：联系人匹配[KLS+17] (X：某应用用户数据库，Y：新用户手机联系人)

研究现状

按照功能分类：

- 隐私集合求交（Private Set Intersection，PSI）：Meadows[Mea86]提出了第一个安全的 PSI 协议；基于不经意传输与不经意伪随机函数[PSZ14，KKRT16，PRTY19，GPR+21，RS21，RR22]。
- 隐私集合求并（Private Set Union，PSU）：基于公钥密码[KS05，Fri07，DC17]；基于对称密码[KRTW19，GMR+21，JSZ+22，ZCL+22]。
- 隐私集合交集计算：电路 PSI[HEK12，PSSZ15，PSTY19]；具有特定功能的隐私集合交集计算协议[DPT20，IKN+20，CZZD22]。
- 非平衡场景：[CLR17，CHLR18，TCLZ22]。

存在问题

尽管目前 PSO 协议的性能已经得到极大提升，但是目前存在如下问题：

- 问题一：作为 PSO 协议底层的关键优化技巧，各类数据结构的使用方式混乱不清、各数据结构之间的效率对比并不清晰。
- 问题二：高效的非平衡场景下隐私集合交集计算的研究较少。

针对问题一展开的研究

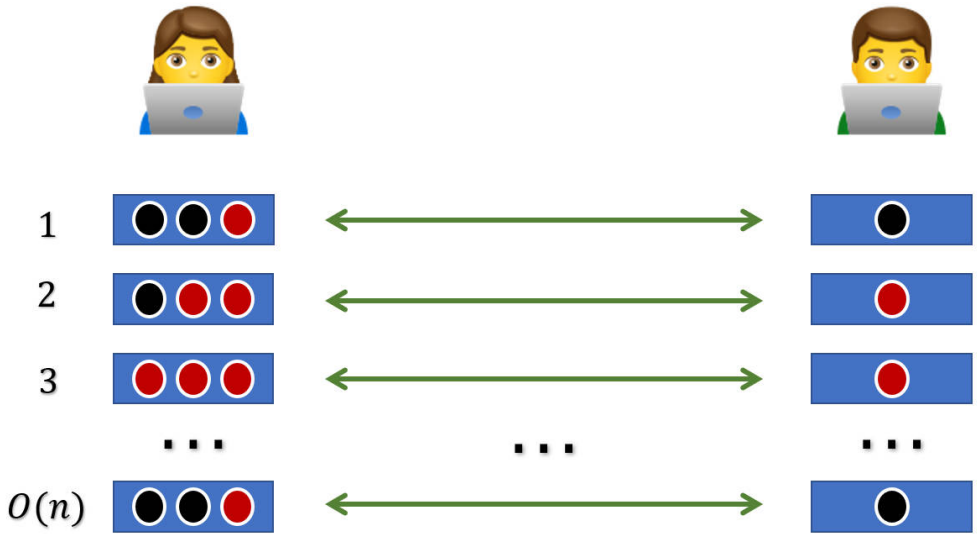
针对问题一：本文将现有 PSO 中常用的数据结构分为三类分别是哈希表、过滤器和不经意键值对存储，并进行了系统的总结。

表 1: PSO 中数据结构的分类

数据结构	分类	作用
哈希表	简单哈希表	数据对齐
	布谷鸟哈希表	
过滤器	布隆过滤器	成员测试
	布谷鸟过滤器	
	真空过滤器	
不经意键值对存储	多项式	数据编码
	乱码布隆过滤器	
	乱码布谷鸟哈希表	

数据对齐：以布谷鸟哈希表为例

- h_1, h_2 将 n 个元素映射到 $m = O(n)$ 个桶中。
- Alice – 简单哈希
  - $x \mapsto h_1(x)$ and $h_2(x)$
  - $B = O(\log n)$
- Bob – 布谷鸟哈希
  - $x \mapsto h_1(x)$ or $h_2(x)$
  - $B = 1$
- 比较次数：$n^2 \to O(n \log n)$



其他数据结构

过滤器、不经意键值对存储

...

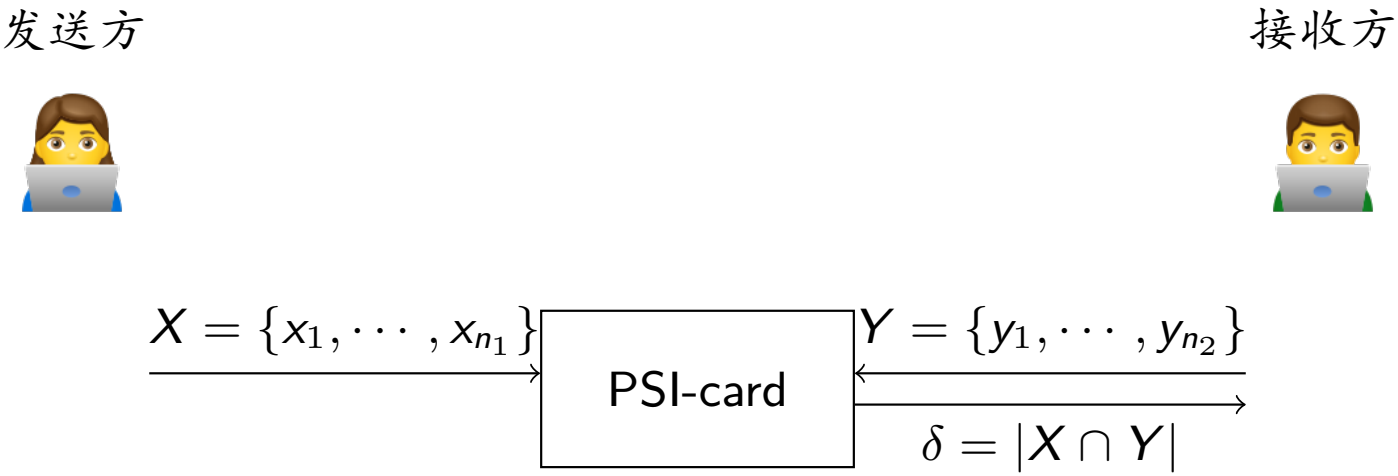
总结其在 PSO 中的应用模式、发挥的主要作用、各类数据结构的性能对比分析与测试。

针对问题二展开的研究

针对问题二：本文提出了非平衡场景下更加高效的具有特定功能的隐私集合交集计算协议，分别为隐私集合求交集势（Private Set Intersection with Cardinality，PSI-card）协议和隐私集合求交集势与和（Private Intersection-Sum-with-Cardinality，PSI-card-sum）协议。
出发点：

- Chen 等人[CLR17]利用全同态加密（Fully Homomorphic Encryption，FHE）构建了通信复杂度与大集合亚线性相关的 PSI 协议。
- Chen 等人[CHLR18]提出了非平衡场景下基于通用两方计算构造的隐私集合交集计算协议。
- Tu 等人[TCLZ22]提出了置换矩阵隐私等值检测协议（Permuted Matrix Private Equality Test，pm-PEQT）并构造了非平衡场景下的 PSU 协议。

PSI-card 定义



CLR17 协议框架

发送方 接收方

$|X| = n_1 \gg |Y| = n_2$

①计算 $F(z) = r \cdot \prod_{x_i \in X}(z - x_i)$

②发送 $[\![y_i]\!] = \mathsf{FHE.Enc}(y_i)$ ←

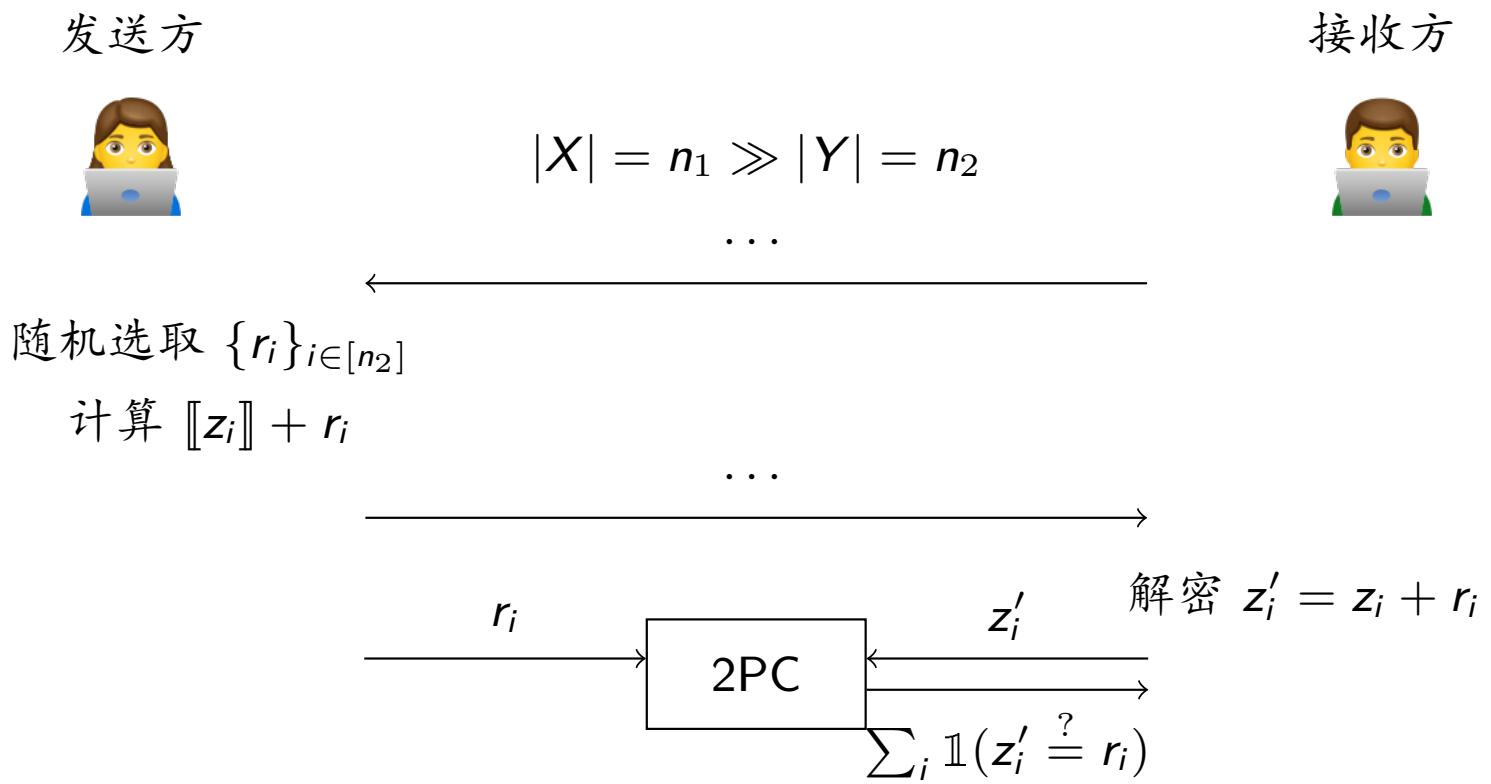
③计算 $[\![y_i^k]\!] \Rightarrow F([\![y_i]\!]) = [\![z_i]\!]$

④发送 $[\![z_i]\!]$ →

解密 $z_i = 0 \Leftrightarrow y_i \in X$

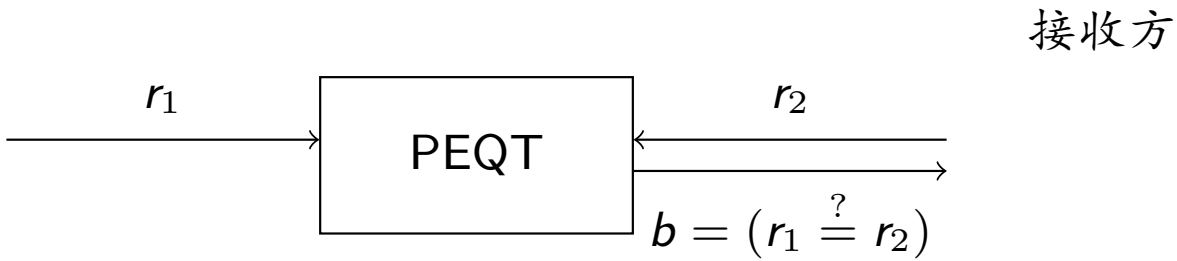
基于 CLR17 协议框架的 PSI-card 协议

在 CLR17 协议框架的基础上，Chen 等人[CHLR18]提出了基于通用两方计算（2PC）构造的 PSI-card 协议。

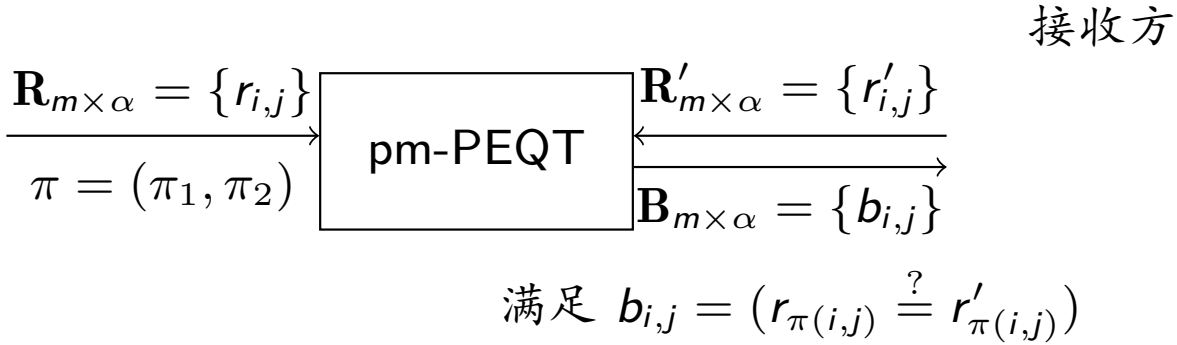


改进的 PSI-card 协议：基于 pm-PEQT 协议

存在问题：基于 2PC 不够高效，且该协议并没附带实现。
改进：利用 pm-PEQT 协议替换 2PC。
PEQT 协议的定义：



pm-PEQT 协议的定义：



改进的 PSI-card 协议：基于 pm-PEQT 协议

发送方  接收方 

$|X| = n_1 \ll |Y| = n_2$

①计算 $F(z) = \prod_{y_i \in Y}(z - y_i)$

②发送 $[\![x_i]\!] = \mathsf{FHE.Enc}(x_i)$ →

③计算 $[\![z_1]\!] + r_1, \cdots, [\![z_{n_1}]\!] + r_{n_1}$

← ④发送 $[\![z_i]\!] + r_i$

解密 $z_i' = z_i + r_i$

$\mathbf{Z}_{n_1} = \{z_i'\}$ → π → pm-PEQT ← $\mathbf{R}_{n_1} = \{r_i\}$

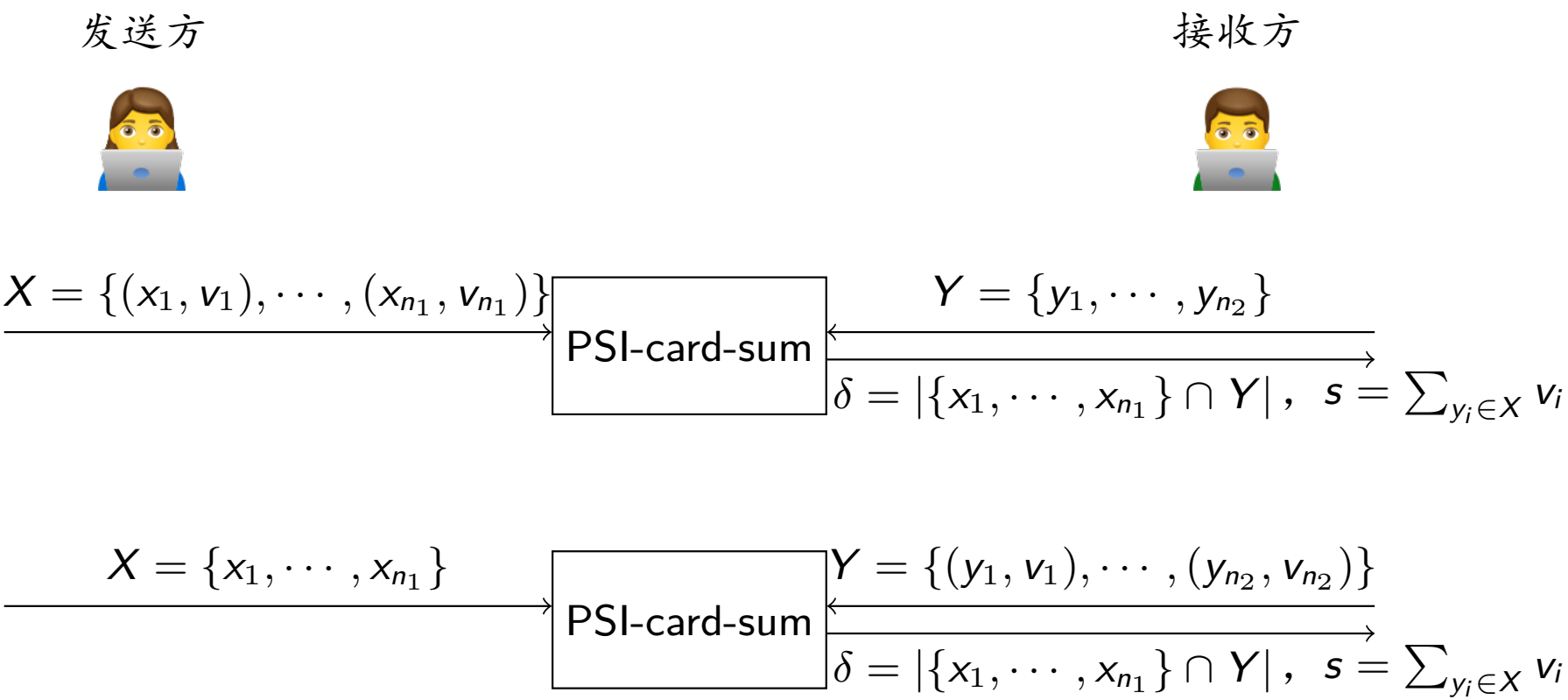
pm-PEQT → $\mathbf{B}_{n_1} = \{b_i\}$

$\delta = \sum_{i=1}^{n_1} b_i$

优化技术

- 布谷鸟哈希表实现数据对齐：桶中元素最多为 $O(\log n)$，降低多项式次数；
- 批处理：密文打包；
- 分窗：发送方只发送一个密文 $[\![x]\!]$，接收方同态计算 $[\![x^k]\!]$ 的乘法电路深度为 $O(\log B)$，发送方额外计算 $[\![x^{2^0}]\!], [\![x^{2^1}]\!], [\![x^{2^2}]\!], \cdots, [\![x^{2^{\log B}}]\!]$，乘法电路深度从 $O(\log B)$ 降低到 $O(\log\log B)$；
- 划分：接收方将每个桶中元素划分成 α 个子集，乘法电路深度进一步降低到 $O(\log\log \frac{B}{\alpha})$。

PSI-card-sum 定义



功能拓展：基于 CLR17 协议框架的 PSI-card-sum 协议

发送方拥有标签值的协议构造：

发送方

接收方

$|X| = n_1 \gg |Y| = n_2$

$\{x_1, \cdots, x_{n_1}\}$ → PSI-card ← Y

PSI-card → $\delta = |\{x_1, \cdots, x_{n_1}\} \cap Y|$

随机选取 $\{e_i\}_{i\in[n_1]}$ 满足 $\sum_{i=1}^{n_1} e_i = 0$

计算 $e_i + v_{\pi(i)}$

$(e_i, e_i + v_{\pi(i)})$ → (2,1)-OT ← b_i

(2,1)-OT → m_i

$s = \sum_{i=1}^{n_1} m_i$

功能拓展：基于 CLR17 协议框架的 PSI-card-sum 协议

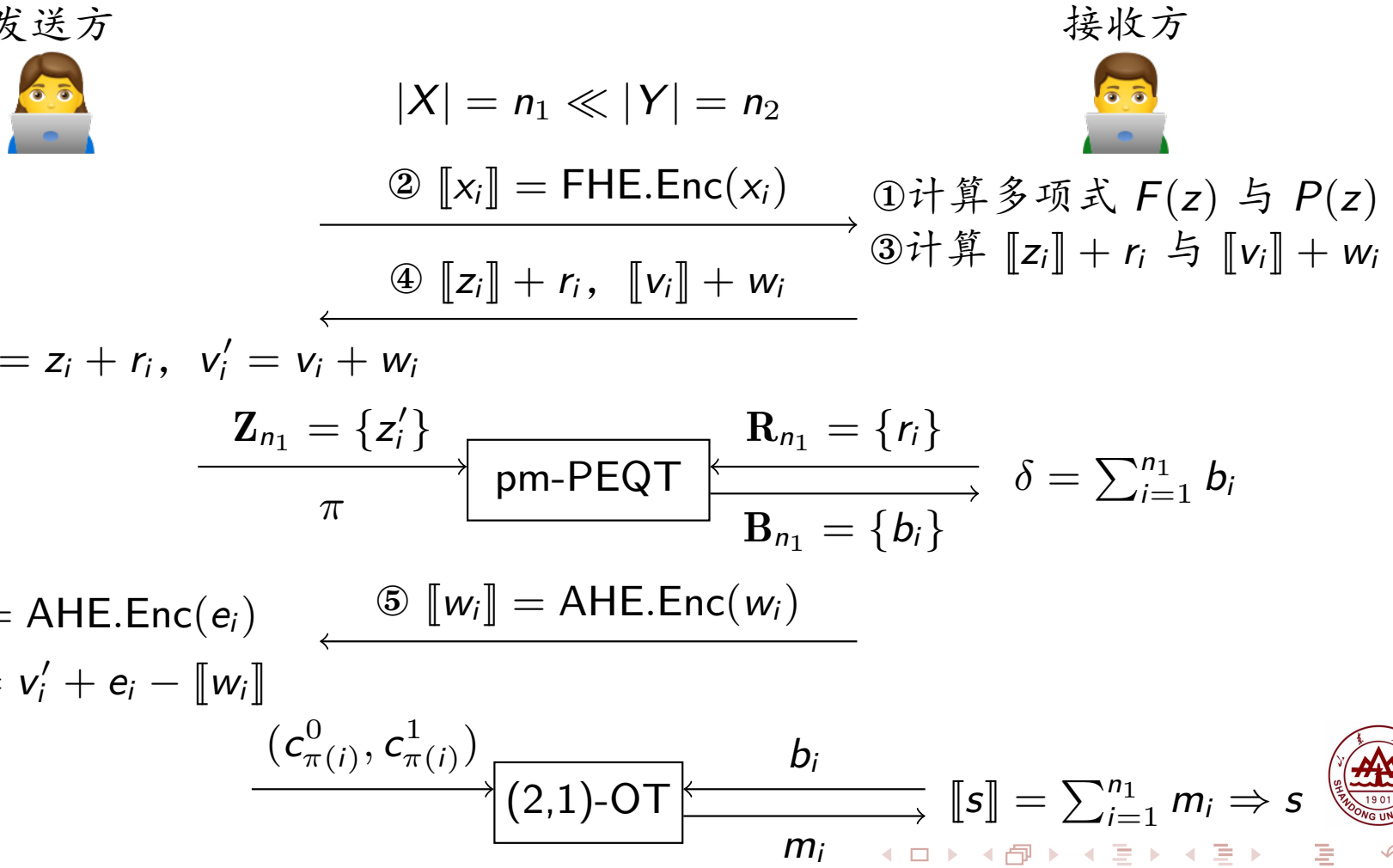
接收方拥有标签值的协议构造：
需解决的问题：$b_i = 1$ 时只能指示 $x_{\pi(i)} \in Y$，无法对应到 Y 中对应的具体标签值 v_i。
解决方法：接收方除计算多项式 $F(z) = \prod_{y_i \in Y}(z - y_i)$ 之外，另外计算多项式 $P(z)$ 满足

$$P(y_i) = v_i$$

容易证明，若 $x \in Y$ 且 $x = y_i$，则 $P(x) = v_i$。

功能拓展：基于 CLR17 协议框架的 PSI-card-sum 协议

接收方拥有标签值的协议构造：



协议实现

本文的协议基于 C++ 实现，并使用如下开源库：

- APSI: https://github.com/microsoft/APSI
- Kunlun: https://github.com/yuchen1024/Kunlun
- libOTe: https://github.com/osu-crypto/libOTe
- paillier-libraries-benchmarks: https://github.com/snipsco/paillier-libraries-benchmarks

实验结果

表 2: PSI-card 协议的时间开销与空间开销（小集合大小 $|X| = 2^{10}$）

$\|Y\|$	协议	空间开销（MB）	时间开销（s）			
			LAN		WAN	
			离线	在线	离线	在线
2^{22}	[CZZD22]	270.4	-	211.2	-	260.6
	本文	4.5	75.4	5.1	77.7	6.3
2^{20}	[CZZD22]	67.6	-	52.5	-	67.9
	本文	2.5	17.6	2.0	19.5	3.2
2^{18}	[CZZD22]	16.9	-	13.2	-	18.1
	本文	2.2	3.6	1.1	5.7	2.1

实验结果

表 3: PSIwS-card-sum 协议的时间开销与空间开销（小集合大小 $|X| = 2^{10}$）

$\|Y\|$	协议	空间开销（MB）	时间开销（s）			
			LAN		WAN	
			离线	在线	离线	在线
2^{22}	[CZZD22]	366.4	-	221	-	287.9
	本文	4.58	75.9	5.4	77.9	7.0
2^{20}	[CZZD22]	91.7	-	55.7	-	73.2
	本文	2.62	17.9	2.2	19.7	3.8
2^{18}	[CZZD22]	22.9	-	17.6	-	22.4
	本文	2.15	3.0	1.1	5.9	2.8

实验结果

表 4: PSIwR-card-sum 协议的时间开销与空间开销（T 表示线程数，小集合大小 $|X| = 2^9$）

$\|Y\|$	协议	空间开销（MB）	时间开销（s）			
			LAN		WAN	
			离线 T=8	在线 T=1	离线 T=8	在线 T=1
2^{22}	[CZZD22]	336.4	-	221	-	287.9
	本文	15.5	232.8	55.3	233.3	77.4
2^{20}	[CZZD22]	91.7	-	55.7	-	73.2
	本文	6.9	45.2	21.2	48.1	30.8
2^{18}	[CZZD22]	22.9	-	17.6	-	22.4
	本文	5.2	6.3	14.2	8.1	21.1

[CHLR18] Hao Chen, Zhicong Huang, Kim Laine, and Peter Rindal.
Labeled psi from fully homomorphic encryption with malicious security.
In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pages 1223–1237, 2018.

[CLR17] Hao Chen, Kim Laine, and Peter Rindal.
Fast private set intersection from homomorphic encryption.
In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 1243–1255, 2017.

[CZZD22] Yu Chen, Min Zhang, Cong Zhang, and Minglang Dong.
Private set operations from multi-query reverse private membership test.
IACR Cryptol. ePrint Arch., page 652, 2022.

[DC17] Alex Davidson and Carlos Cid.
An efficient toolkit for computing private set operations.
In *Australasian Conference on Information Security and Privacy*, pages 261–278. Springer, 2017.

[DPT20] Thai Duong, Duong Hieu Phan, and Ni Trieu.
Catalic: Delegated psi cardinality with applications to contact tracing.
In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 870–899. Springer, 2020.

[Fri07] Keith Frikken.
Privacy-preserving set union.
In *International Conference on Applied Cryptography and Network Security*, pages 237–252. Springer, 2007.

[GMR+21] Gayathri Garimella, Payman Mohassel, Mike Rosulek, Saeed Sadeghian, and Jaspal Singh.
Private set operations from oblivious switching.
In Juan A. Garay, editor, *Public-Key Cryptography - PKC 2021 - 24th IACR International Conference on Practice and Theory of Public Key Cryptography, Virtual Event, May 10-13, 2021, Proceedings, Part II*, volume 12711 of *Lecture Notes in Computer Science*, pages 591–617. Springer, 2021.

[GPR+21] Gayathri Garimella, Benny Pinkas, Mike Rosulek, Ni Trieu, and Avishay Yanai.
Oblivious key-value stores and amplification for private set intersection.
In *Annual International Cryptology Conference*, pages 395–425. Springer, 2021.

[HEK12] Yan Huang, David Evans, and Jonathan Katz.
Private set intersection: Are garbled circuits better than custom protocols?
In *19th Annual Network and Distributed System Security Symposium, NDSS 2012, San Diego, California, USA, February 5-8, 2012*. The Internet Society, 2012.

[IKN+20] Mihaela Ion, Ben Kreuter, Ahmet Erhan Nergiz, Sarvar Patel, Shobhit Saxena, Karn Seth, Mariana Raykova, David Shanahan, and Moti Yung.
On deploying secure computing: Private intersection-sum-with-cardinality.
In *IEEE European Symposium on Security and Privacy, EuroS&P 2020, Genoa, Italy, September 7-11, 2020*, pages 370–389. IEEE, 2020.

[JSZ+22] Yanxue Jia, Shifeng Sun, Hong-Sheng Zhou, Jiajun Du, and Dawu Gu.
Shuffle-based private set union: Faster and more secure.
In Kevin R. B. Butler and Kurt Thomas, editors, *31st USENIX Security Symposium, USENIX Security 2022, Boston, MA, USA, August 10-12, 2022*, pages 2947–2964. USENIX Association, 2022.

[KKRT16] Vladimir Kolesnikov, Ranjit Kumaresan, Mike Rosulek, and Ni Trieu.
Efficient batched oblivious prf with applications to private set intersection.
In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 818–829, 2016.

[KLS+17] Ágnes Kiss, Jian Liu, Thomas Schneider, N Asokan, and Benny Pinkas.
Private set intersection for unequal set sizes with mobile applications.
Proc. Priv. Enhancing Technol., 2017(4):177–197, 2017.

[KRTW19] Vladimir Kolesnikov, Mike Rosulek, Ni Trieu, and Xiao Wang.
Scalable private set union from symmetric-key techniques.
In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 636–666. Springer, 2019.

[KS05] Lea Kissner and Dawn Song.
Privacy-preserving set operations.
In *Annual International Cryptology Conference*, pages 241–257. Springer, 2005.

[Mea86] Catherine A. Meadows.
A more efficient cryptographic matchmaking protocol for use in the absence of a continuously available third party.
In *Proceedings of the 1986 IEEE Symposium on Security and Privacy, Oakland, California, USA, April 7-9, 1986*, pages 134–137. IEEE Computer Society, 1986.

[PRTY19] Benny Pinkas, Mike Rosulek, Ni Trieu, and Avishay Yanai.
Spot-light: lightweight private set intersection from sparse ot extension.
In *Annual International Cryptology Conference*, pages 401–431. Springer, 2019.

[PSSZ15] Benny Pinkas, Thomas Schneider, Gil Segev, and Michael Zohner.
Phasing: Private set intersection using permutation-based hashing.
In Jaeyeon Jung and Thorsten Holz, editors, *24th USENIX Security Symposium, USENIX Security 15, Washington, D.C., USA, August 12-14, 2015*, pages 515–530. USENIX Association, 2015.

[PSTY19] Benny Pinkas, Thomas Schneider, Oleksandr Tkachenko, and Avishay Yanai.
Efficient circuit-based psi with linear communication.
In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 122–153. Springer, 2019.

[PSZ14] Benny Pinkas, Thomas Schneider, and Michael Zohner.
Faster private set intersection based on {OT} extension.
In *23rd USENIX Security Symposium (USENIX Security 14)*, pages 797–812, 2014.

[RR22] Srinivasan Raghuraman and Peter Rindal.
Blazing fast psi from improved okvs and subfield vole.
Cryptology ePrint Archive, Paper 2022/320, 2022.
https://eprint.iacr.org/2022/320.

[RS21] Peter Rindal and Phillipp Schoppmann.
Vole-psi: fast oprf and circuit-psi from vector-ole.
In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 901–930. Springer, 2021.

[TCLZ22] Binbin Tu, Yu Chen, Qi Liu, and Cong Zhang.
Fast unbalanced private set union from fully homomorphic encryption.
Cryptology ePrint Archive, Paper 2022/653, 2022.
https://eprint.iacr.org/2022/653.

[ZCL+22] Cong Zhang, Yu Chen, Weiran Liu, Min Zhang, and Dongdai Lin.
Optimal private set union from multi-query reverse private membership test.
Cryptology ePrint Archive, 2022.

Thanks!