隐私集合计算中的关键技术研究

毕业论文开题报告

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）

2022 年 6 月 15 日

1 课题背景

2 研究现状

3 研究内容

4 研究思路

5 进度安排

6 参考文献

1 课题背景
2 研究现状
3 研究内容
4 研究思路
5 进度安排
6 参考文献

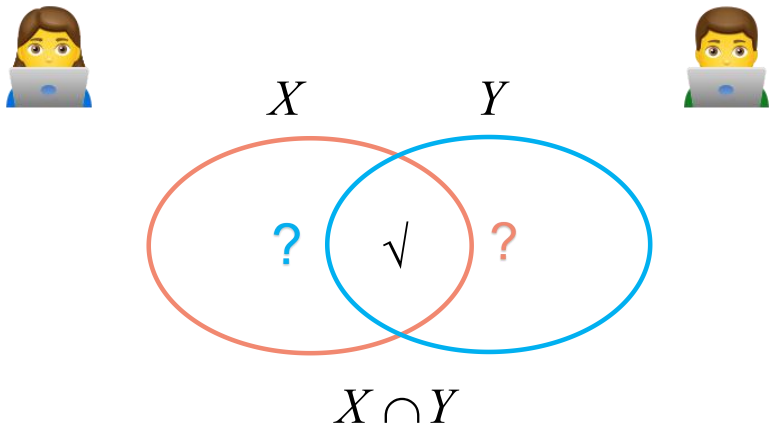
- 数据价值
  - 大数据时代，海量数据的交叉计算可以为科研、医疗、金融等领域提供更好的支持；
  - 数据已成为社会生产的新要素，是国家基础性战略资源。

- 数据价值
  - 大数据时代，海量数据的交叉计算可以为科研、医疗、金融等领域提供更好的支持；
  - 数据已成为社会生产的新要素，是国家基础性战略资源。
- 安全需求
  - 数据安全和隐私泄漏问题频发；
  - 《数据安全法》、《个人信息保护法》等法律法规的颁布实施。

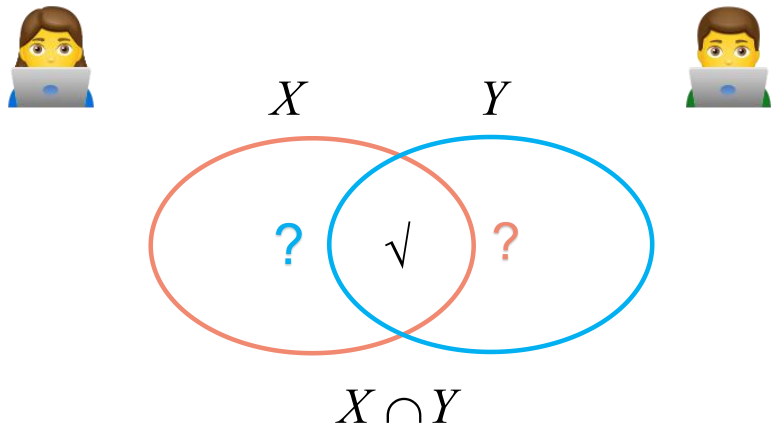
- 数据价值
  - 大数据时代，海量数据的交叉计算可以为科研、医疗、金融等领域提供更好的支持；
  - 数据已成为社会生产的新要素，是国家基础性战略资源。
- 安全需求
  - 数据安全和隐私泄漏问题频发；
  - 《数据安全法》、《个人信息保护法》等法律法规的颁布实施。
- 既要应用数据，又要保护数据安全 $\Rightarrow$ 安全多方计算

- 隐私集合计算（Private Set Operation，PSO）属于安全多方计算领域的热点问题，包括隐私集合求交（Private Set Intersection，PSI）和隐私集合求并（Private Set Union，PSU）。

- 隐私集合计算（Private Set Operation，PSO）属于安全多方计算领域的热点问题，包括隐私集合求交（Private Set Intersection，PSI）和隐私集合求并（Private Set Union，PSU）。
- 以 PSI 为例：



- 隐私集合计算（Private Set Operation，PSO）属于安全多方计算领域的热点问题，包括隐私集合求交（Private Set Intersection，PSI）和隐私集合求并（Private Set Union，PSU）。
- 以 PSI 为例：



- 应用场景：联系人匹配[KLS+17] (X :Whatsapp 用户数据库，Y：新用户手机联系人)
计算广告转化率[IKN+17] (X：浏览广告的用户，Y：购买相应商品的用户)

1 课题背景

2 研究现状

3 研究内容

4 研究思路

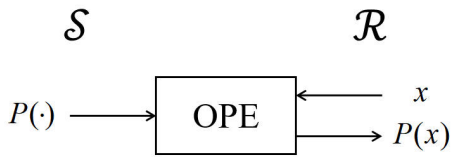
5 进度安排

6 参考文献

PSI 协议

根据底层所采用技术的不同，PSI 协议可以分为：

- 基于不经意多项式求值（OPE）的 PSI 协议

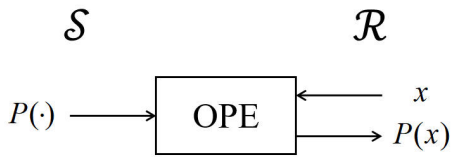


  - 构造：同态加密[FNP04]

PSI 协议

根据底层所采用技术的不同，PSI 协议可以分为：

- 基于不经意多项式求值（OPE）的 PSI 协议

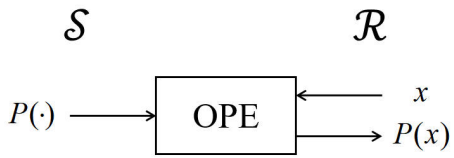


  - 构造：同态加密[FNP04]
- 基于布隆过滤器（BF）的 PSI 协议
  - [DCW13, RR17]
  - PSI 协议能处理的集合数量首次突破了亿级别

PSI 协议

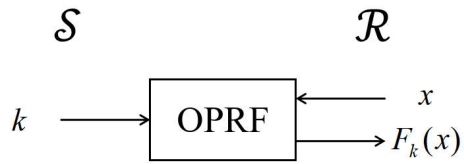
根据底层所采用技术的不同，PSI 协议可以分为：

- 基于不经意多项式求值（OPE）的 PSI 协议



  - 构造：同态加密[FNP04]
- 基于布隆过滤器（BF）的 PSI 协议
  - [DCW13, RR17]
  - PSI 协议能处理的集合数量首次突破了亿级别

- 基于不经意伪随机函数（OPRF）的 PSI 协议



  - 构造：Diffie-Hellman 密钥交换[Mea86]、Blind-RSA[DCT09]、不经意传输（OT）[PSZ14, PSSZ15, KKRT16, PRTY19, CM20]、不经意向量线性求值（Vector-OLE）[RS21]

PSU 协议

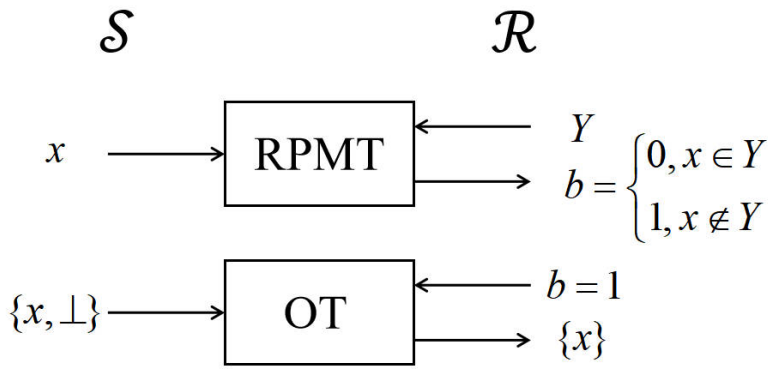
根据底层所采用技术的不同，PSU 协议可以分为：

- 基于公钥加密体系的 PSU 协议
  - 利用 AHE 加密接收方集合的表示 (多项式或布隆过滤器)，并对密文执行大量计算[KS05, Fri07, DC17]

PSU 协议

根据底层所采用技术的不同，PSU 协议可以分为：

- 基于公钥加密体系的 PSU 协议
  - 利用 AHE 加密接收方集合的表示 (多项式或布隆过滤器)，并对密文执行大量计算[KS05, Fri07, DC17]
- 基于对称加密体系的 PSU 协议



- 基于 OT 扩展构造 RPMT 子协议[KRTW19, ZCL+22]

1 课题背景

2 研究现状

3 研究内容

4 研究思路

5 进度安排

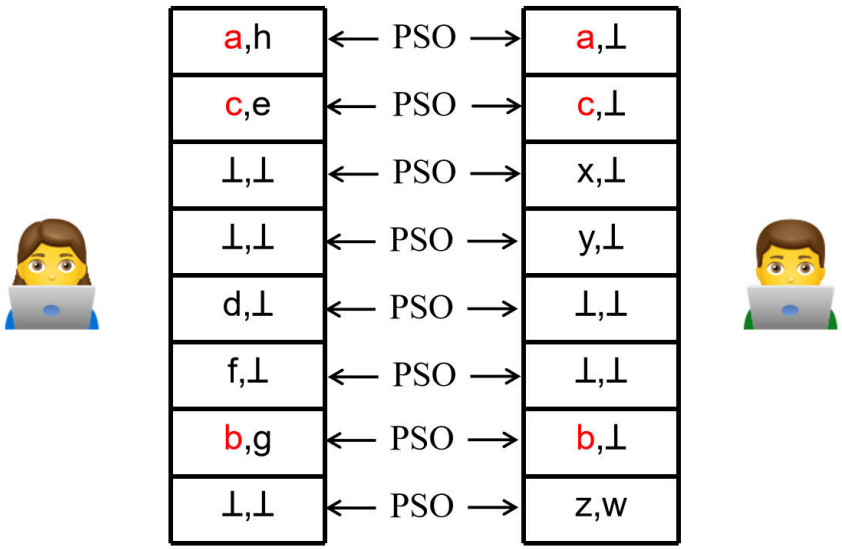
6 参考文献

PSO 协议中的数据结构

- 在 PSO 协议的设计中，往往运用了一些高级的数据结构，对于降低协议的渐进复杂度起到了重要的作用。

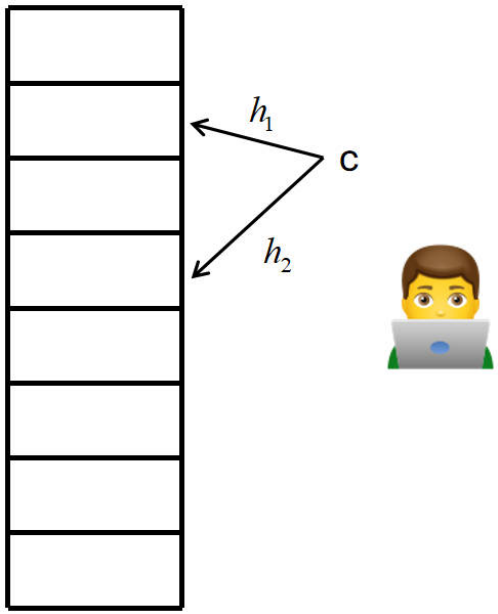
数据结构	类型	作用
哈希表	简单哈希表	数据对齐
	置换哈希表	
	布谷鸟哈希表	
过滤器	布隆过滤器（BF）	成员测试
	布谷鸟过滤器（CF）	
不经意键值对存储	乱码布隆过滤器（GBF）	数据编码
	乱码布谷鸟哈希表（3H-GCT）	

哈希表：数据对齐



- 发送方和接收方将各自集合中的元素分别映射到两个哈希表中，每个哈希表有 B 个桶
  - 两方相同的元素会被映射到相同索引的桶中
  - 填充哑元以防止额外信息泄露
- 逐桶执行 PSI 或 PSU 子协议

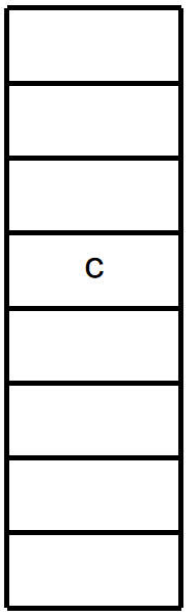
布谷鸟哈希表



选择两个随机的哈希函数 h_1 和 h_2

- 如果 $h_1(x)$ 或 $h_2(x)$ 为空桶
  - 将元素 x 放入桶中

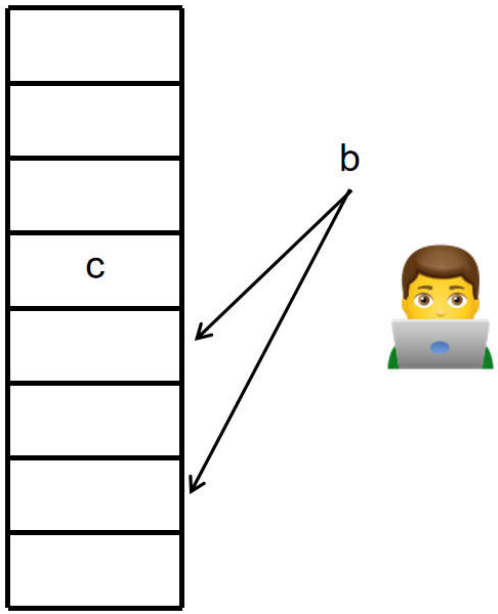
布谷鸟哈希表



选择两个随机的哈希函数 h_1 和 h_2

- 如果 $h_1(x)$ 或 $h_2(x)$ 为空桶
  - 将元素 x 放入桶中

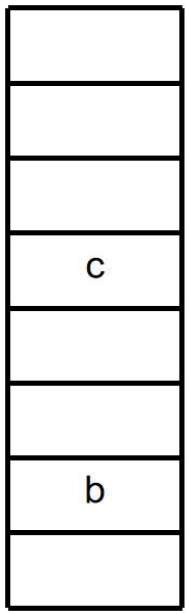
布谷鸟哈希表



选择两个随机的哈希函数 h_1 和 h_2

- 如果 $h_1(x)$ 或 $h_2(x)$ 为空桶
  - 将元素 x 放入桶中

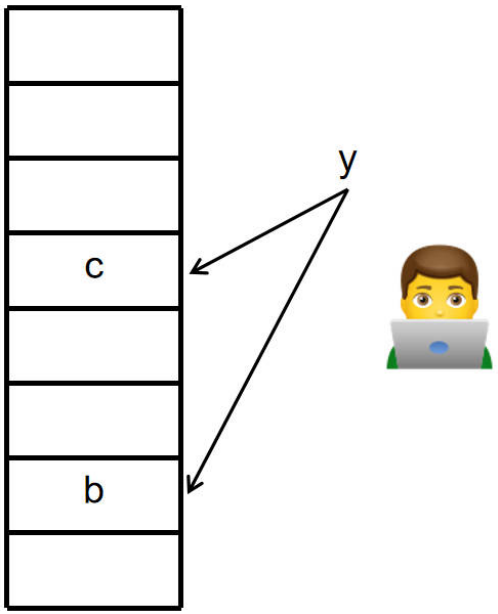
布谷鸟哈希表



选择两个随机的哈希函数 h_1 和 h_2

- 如果 $h_1(x)$ 或 $h_2(x)$ 为空桶
  - 将元素 x 放入桶中

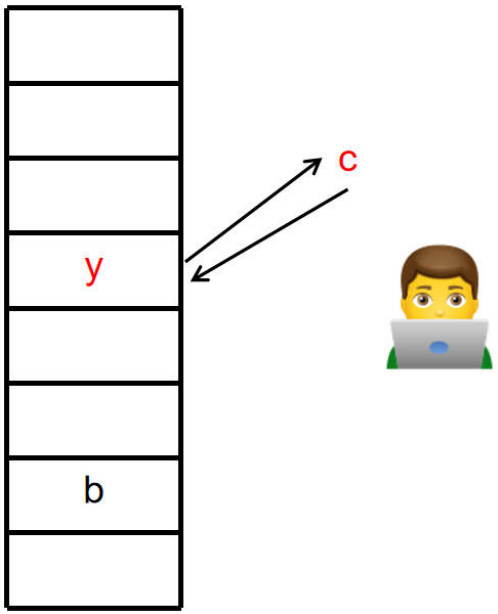
布谷鸟哈希表



选择两个随机的哈希函数 h_1 和 h_2

- 如果 $h_1(x)$ 或 $h_2(x)$ 为空桶
  - 将元素 x 放入桶中
- 如果 $h_1(x)$ 和 $h_2(x)$ 都不为空
  - 随机选择其中一个元素 x' 逐出，并对 x' 进行递归操作

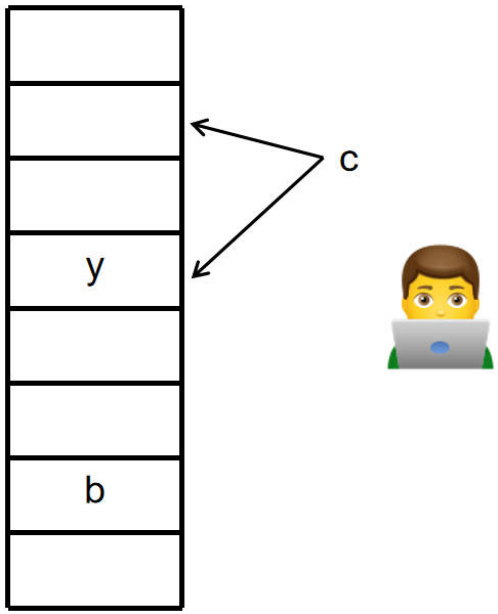
布谷鸟哈希表



选择两个随机的哈希函数 h_1 和 h_2

- 如果 $h_1(x)$ 或 $h_2(x)$ 为空桶
  - 将元素 x 放入桶中
- 如果 $h_1(x)$ 和 $h_2(x)$ 都不为空
  - 随机选择其中一个元素 x' 逐出，并对 x' 进行递归操作

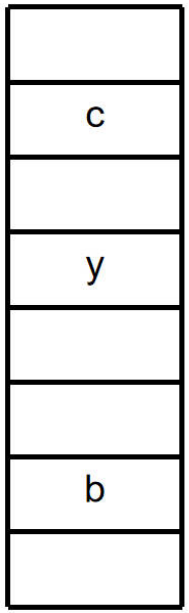
布谷鸟哈希表



选择两个随机的哈希函数 h_1 和 h_2

- 如果 $h_1(x)$ 或 $h_2(x)$ 为空桶
  - 将元素 x 放入桶中
- 如果 $h_1(x)$ 和 $h_2(x)$ 都不为空
  - 随机选择其中一个元素 x' 逐出，并对 x' 进行递归操作

布谷鸟哈希表

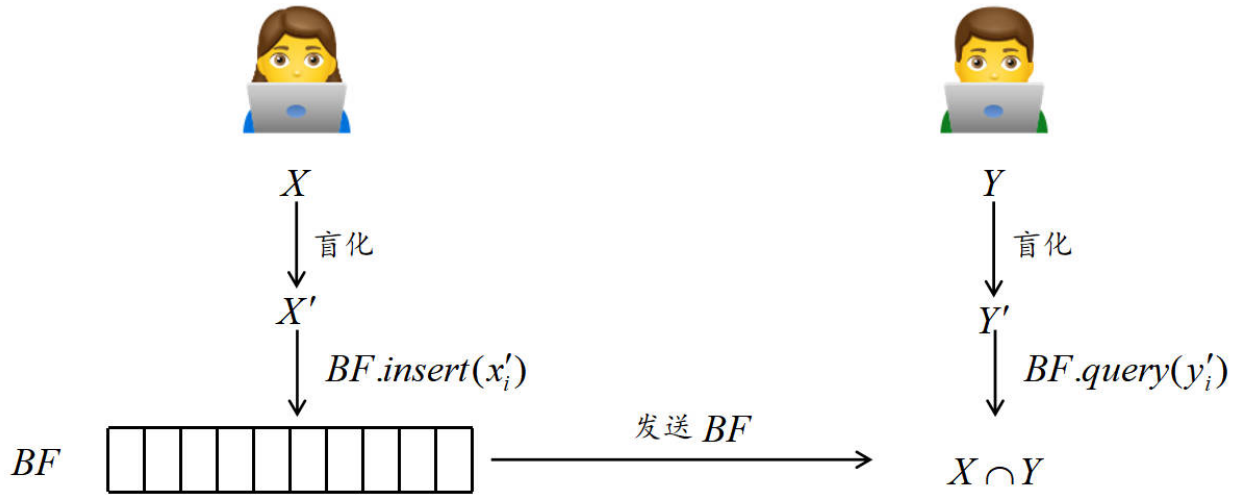


选择两个随机的哈希函数 h_1 和 h_2

- 如果 $h_1(x)$ 或 $h_2(x)$ 为空桶
  - 将元素 x 放入桶中
- 如果 $h_1(x)$ 和 $h_2(x)$ 都不为空
  - 随机选择其中一个元素 x' 逐出，并对 x' 进行递归操作

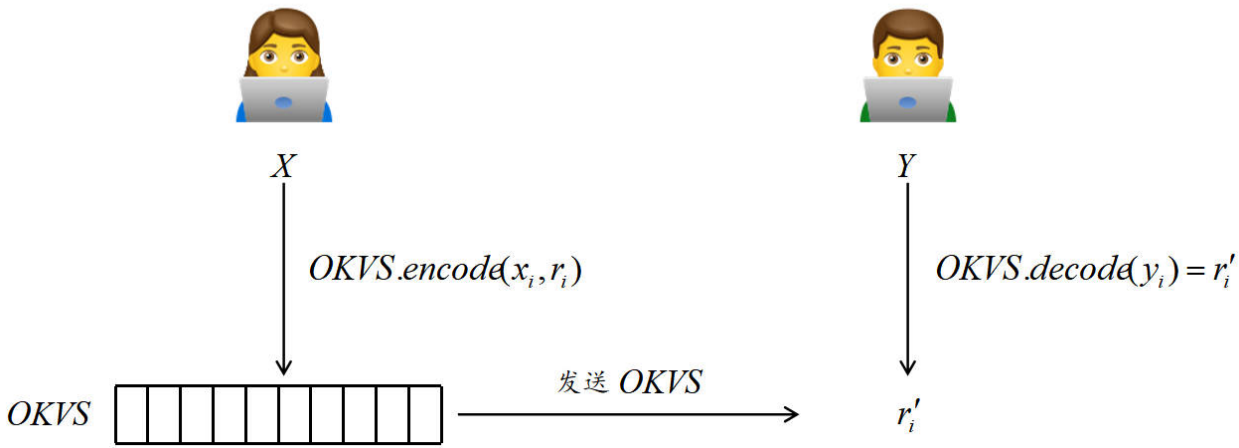
过滤器：成员测试

- 一种 PSI 协议的设计范式可以概括为两个参与方将各自集合中的元素盲化后直接对比得到交集，如基于不经意伪随机函数的 PSI 协议
- 利用布隆过滤器或布谷鸟过滤器进行成员测试可以降低通信开销



不经意键值对存储：数据编码

- 一些 PSI 和 PSU 协议中利用了多项式或乱码布隆过滤器对集合中的元素进行编码
- Garimella 等人将上述技术抽象成了不经意键值对存储（OKVS）
- 通过构造效率更高的 OKVS 实例化（如，3H-GCT）可以直接得到更高效的 PSI 和 PSU 协议



1 课题背景

2 研究现状

3 研究内容

4 研究思路

5 进度安排

6 参考文献

- 性能分析
  - 失败概率分析，如何选取合适的参数包括哈希函数的个数和桶的个数等使得各类数据结构构造时失败概率足够小，如2^{-40}。

- 性能分析
  - 失败概率分析，如何选取合适的参数包括哈希函数的个数和桶的个数等使得各类数据结构构造时失败概率足够小，如 2^{-40}。
  - 构造效率分析，各类数据结构是否可以并行插入元素。

- 性能分析
  - 失败概率分析，如何选取合适的参数包括哈希函数的个数和桶的个数等使得各类数据结构构造时失败概率足够小，如 2^{-40}。
  - 构造效率分析，各类数据结构是否可以并行插入元素。
  - 空间效率分析，各类数据结构插入相同元素个数时表的大小。

- 性能分析
  - 失败概率分析，如何选取合适的参数包括哈希函数的个数和桶的个数等使得各类数据结构构造时失败概率足够小，如 2^{-40}。
  - 构造效率分析，各类数据结构是否可以并行插入元素。
  - 空间效率分析，各类数据结构插入相同元素个数时表的大小。
- 基准测试
  - 关于不同哈希函数的调研。可供选择的哈希函数包括 Murmur 哈希和 Bob 哈希等。

- 性能分析
  - 失败概率分析，如何选取合适的参数包括哈希函数的个数和桶的个数等使得各类数据结构构造时失败概率足够小，如 2^{-40}。
  - 构造效率分析，各类数据结构是否可以并行插入元素。
  - 空间效率分析，各类数据结构插入相同元素个数时表的大小。
- 基准测试
  - 关于不同哈希函数的调研。可供选择的哈希函数包括 Murmur 哈希和 Bob 哈希等。
  - 关于 3H-GCT 的代码实现。3H-GCT 与其他数据结构不同，构造时元素需以特定的顺序插入，即需要处理遇到“环”的情况保证成功率。

- 性能分析
  - 失败概率分析，如何选取合适的参数包括哈希函数的个数和桶的个数等使得各类数据结构构造时失败概率足够小，如 2^{-40}。
  - 构造效率分析，各类数据结构是否可以并行插入元素。
  - 空间效率分析，各类数据结构插入相同元素个数时表的大小。
- 基准测试
  - 关于不同哈希函数的调研。可供选择的哈希函数包括 Murmur 哈希和 Bob 哈希等。
  - 关于 3H-GCT 的代码实现。3H-GCT 与其他数据结构不同，构造时元素需以特定的顺序插入，即需要处理遇到“环”的情况保证成功率。
  - 函数接口的设计以及各种数据结构产生的类之间是否存在继承与派生关系。

1 课题背景
2 研究现状
3 研究内容
4 研究思路
5 进度安排
6 参考文献

研究进度及具体时间安排		
起讫日期	主要研究内容	预期结果
2022.04~2022.06	查阅国内外文献，完成文献综述与开题报告	完成开题报告
2022.06~2022.10	各类数据结构的代码实现与对比分析	代码实现
2022.10~2023.02	撰写毕业论文，完成初稿	毕业论文初稿
2023.02~2023.04	结合导师意见不断完善论文	毕业论文定稿
2023.04~2023.06	进行论文答辩前的相关准备	通过答辩

1 课题背景

2 研究现状

3 研究内容

4 研究思路

5 进度安排

6 参考文献

[CM20] Melissa Chase and Peihan Miao.
Private set intersection in the internet setting from lightweight oblivious prf.
In *Annual International Cryptology Conference*, pages 34–63. Springer, 2020.

[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.

[DCT09] Emiliano De Cristofaro and Gene Tsudik.
Practical private set intersection protocols with linear computational and bandwidth complexity.
Cryptology ePrint Archive, 2009.

[DCW13] Changyu Dong, Liqun Chen, and Zikai Wen.
When private set intersection meets big data: an efficient and scalable protocol.
In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pages 789–800, 2013.

[FNP04] Michael J Freedman, Kobbi Nissim, and Benny Pinkas.
Efficient private matching and set intersection.
In *International conference on the theory and applications of cryptographic techniques*, pages 1–19. Springer, 2004.

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

[IKN+17] Mihaela Ion, Ben Kreuter, Erhan Nergiz, Sarvar Patel, Shobhit Saxena, Karn Seth, David Shanahan, and Moti Yung.
Private intersection-sum protocol with applications to attributing aggregate ad conversions.
Cryptology ePrint Archive, 2017.

[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 Meadows.
A more efficient cryptographic matchmaking protocol for use in the absence of a continuously available third party.
In *1986 IEEE Symposium on Security and Privacy*, pages 134–134. IEEE, 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 *24th USENIX Security Symposium (USENIX Security 15)*, pages 515–530, 2015.

[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.

[RR17] Peter Rindal and Mike Rosulek.
Improved private set intersection against malicious adversaries.
In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 235–259. Springer, 2017.

[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.

[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!