Fuzzy Private Set Intersection for Real-World Datasets

Satvinder Singh

PURDUE UNIVERSITY

(Joint work with Yanxue Jia, Aniket Kate)

2026-October-27 -- NIST Workshop on Privacy-Enhancing Cryptography 2026

What is Fuzzy PSI?



{John Smith,
Daniel O'Connor,
Elena Rostova,
Maya Sterling}



{Olivia Chen,
John Smith,
Benjamin Brooks,
Daniel O'Connor}

Output: {John Smith, Daniel O'Connor}

Standard Private Set Intersection



{John A. Smith,
Daniel OConnor,
Elena Rostova,
Maya Sterling}



{Olivia Chen,
John Smith,
Benjamin Brooks,
Daniël O'Connor}

Output: {John A. Smith, Daniël O'Connor}

Fuzzy Private Set Intersection

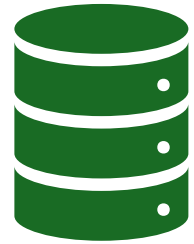
Why do we care?



Ride Sharing



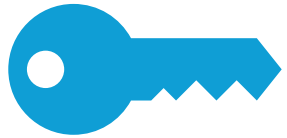
Biometric Authentication



Database Joins



Image Matching

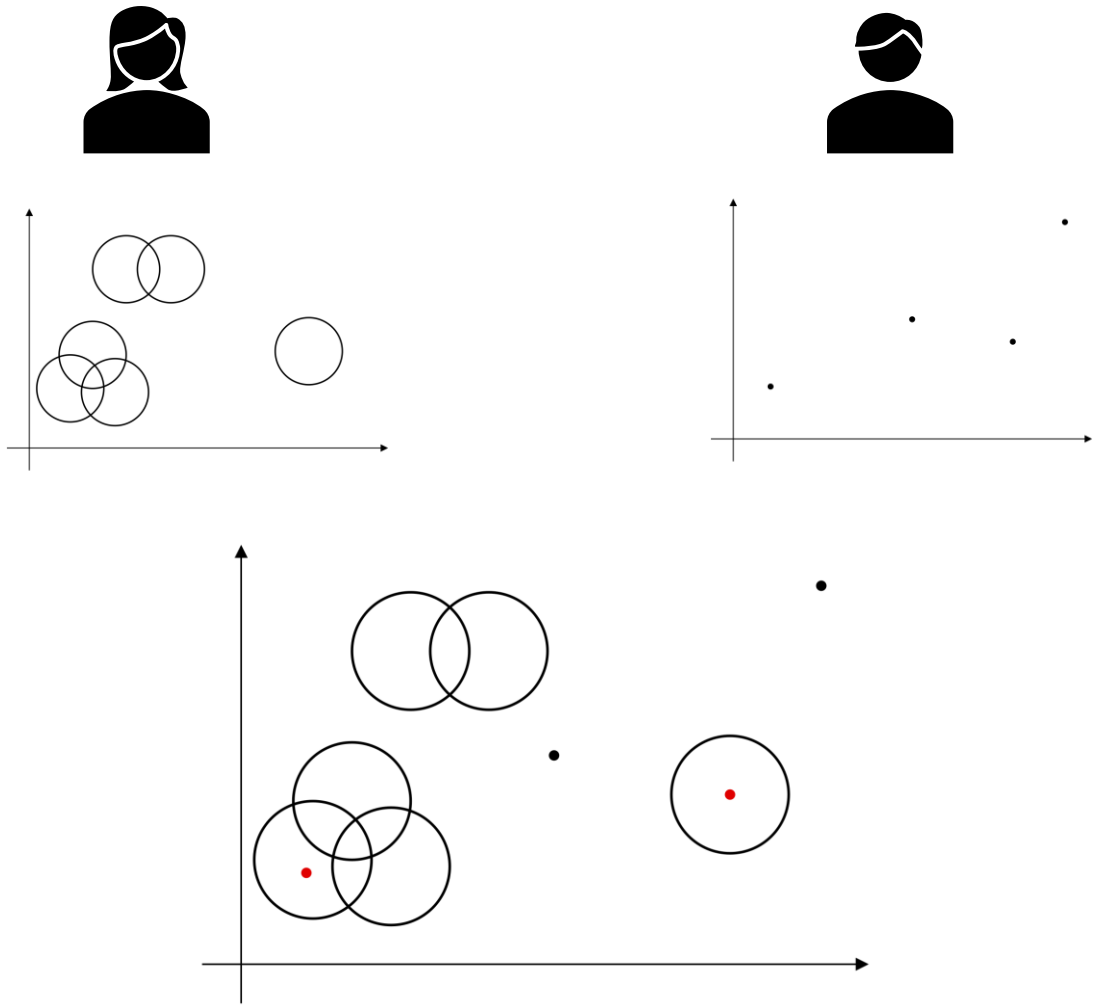


Password Matching

Progress so far

- Lots of papers: [UCKBL'21], [GRS'22], [GRS'23], [CFR'23], [GGM'24], [BP'24], [GQLLW'24], [RRX'24], [PSTKZ'25], [BGMP'25], [BP'25], [DZL'25], [ZCCBLLWW'25], [BN'26], [YHJWWZ'26], [YHWDWZ'26], [SJK'26], …
- What's the ***best*** protocol?

A geometric interpretation of Fuzzy PSI



Parameters

m: number of circles/balls

n: number of points

d: dimension

δ: radius of circles

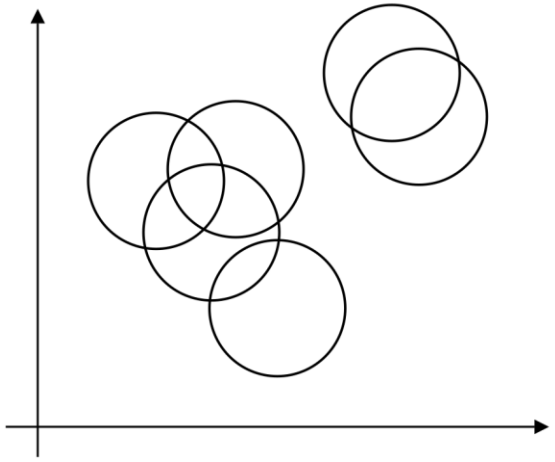
l_p: norm

"Best" protocol:

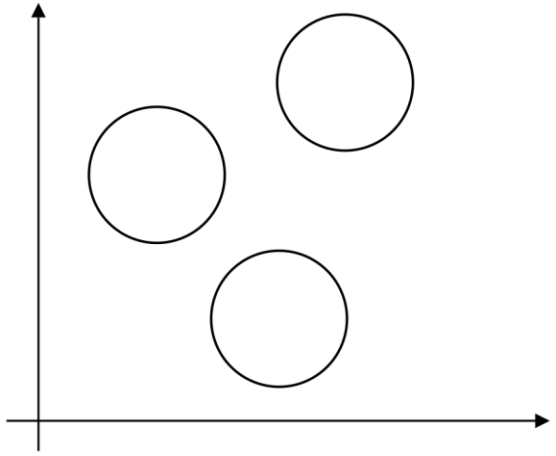
$O((\delta d + p\log\delta)m + \delta dn)$

Are we done?

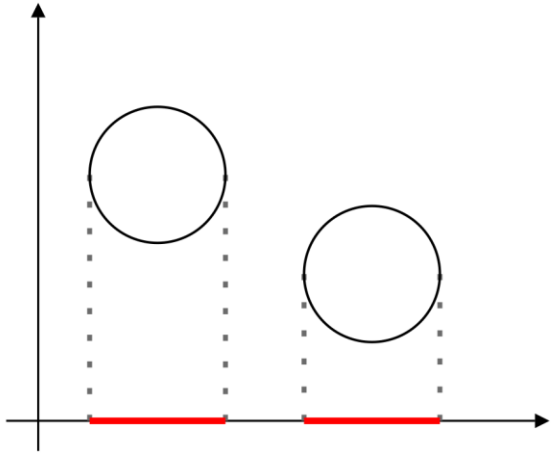
Assumptions about underlying inputs



Real-World



Disjoint Set

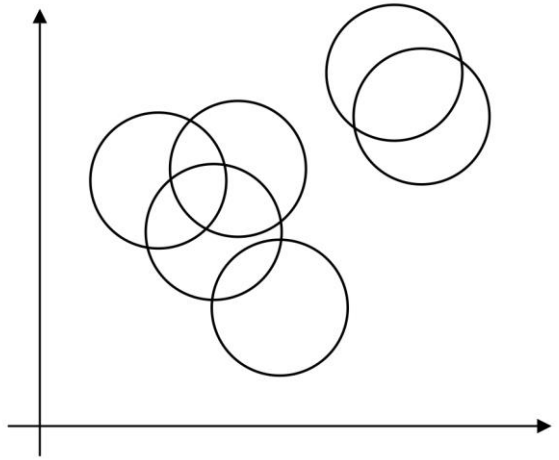


Disjoint Projections

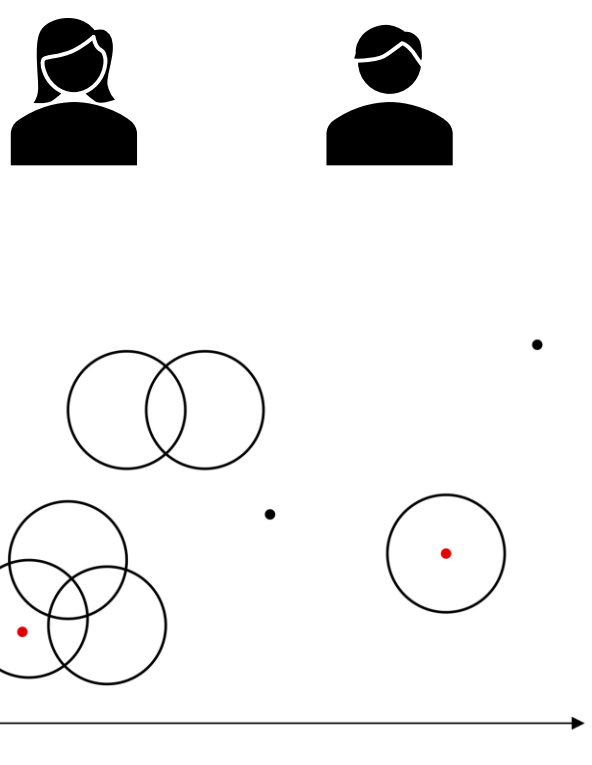
Can we salvage all the cool protocols?

A notion of density

- C : maximum number of circles that overlap at any point
- Disjoint set assumption: $C = 0$
- Observation: C is much smaller than the total number of circles (C = 19k when m=1Mil for NYC taxi trip locations)

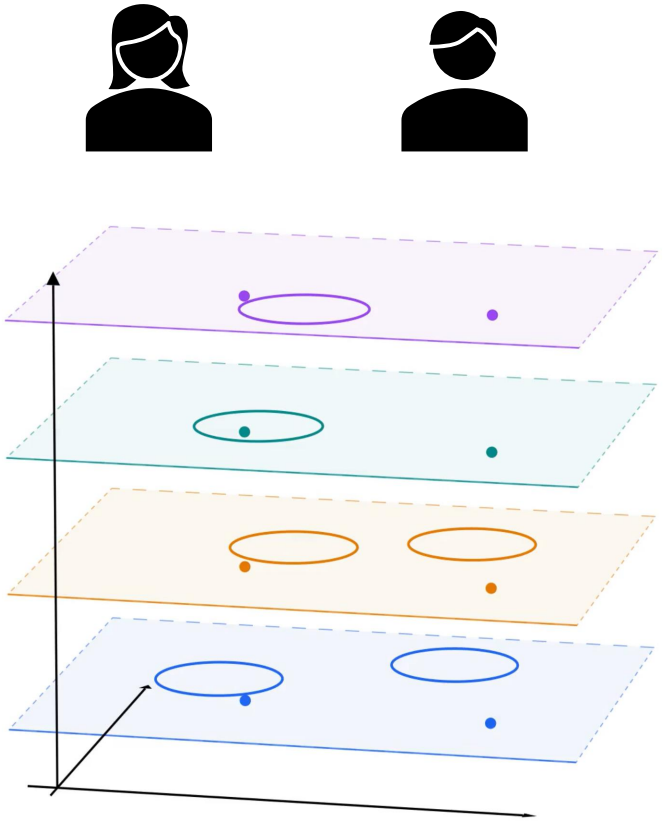


Our Approach

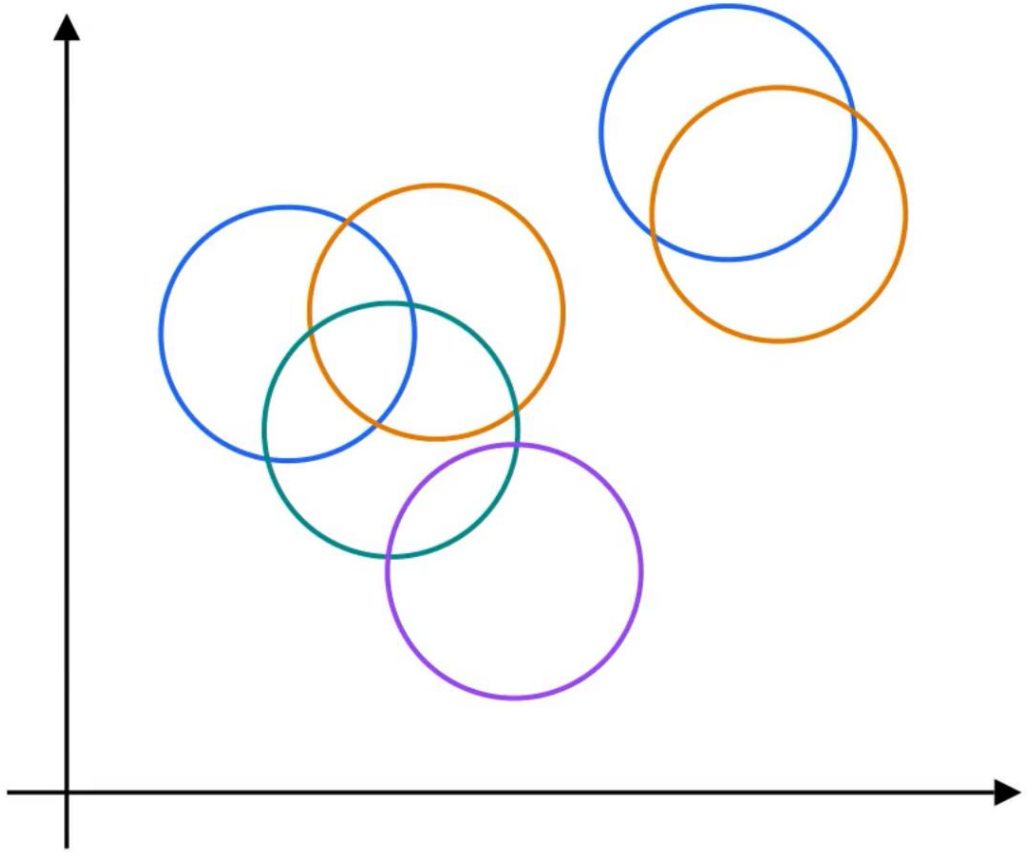


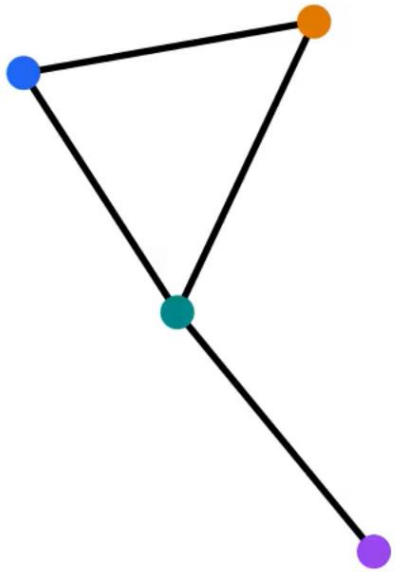
Our Transformation

1. Circles become disjoint
2. Intersection is preserved



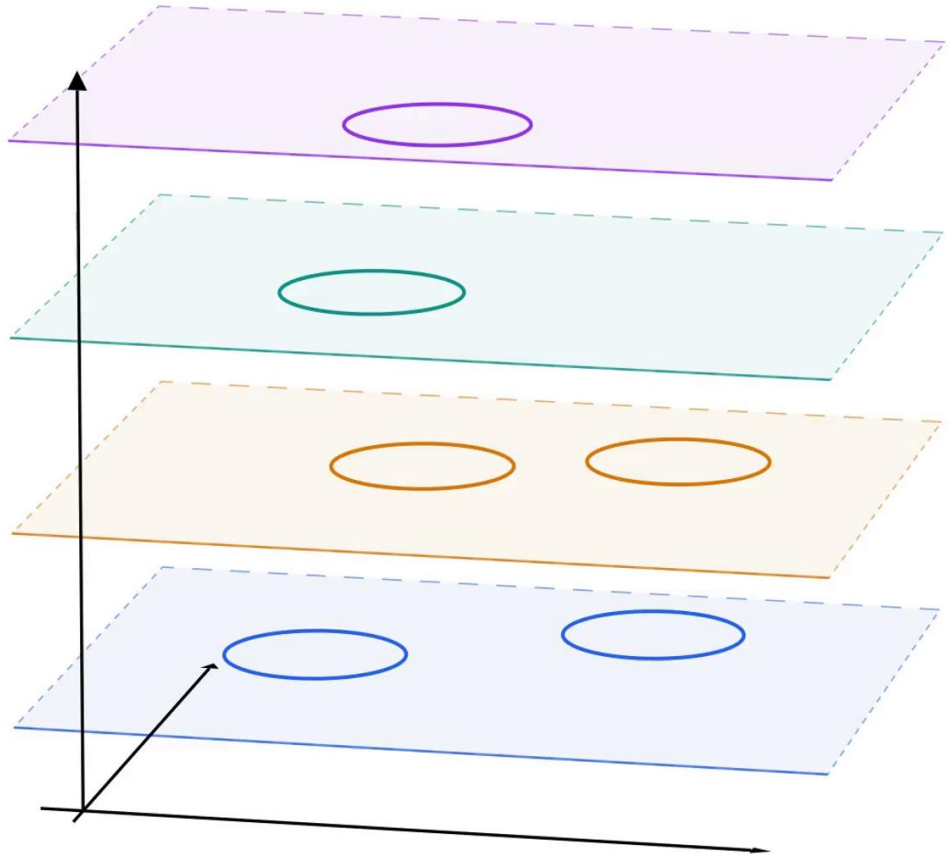
Assign colors via graph coloring



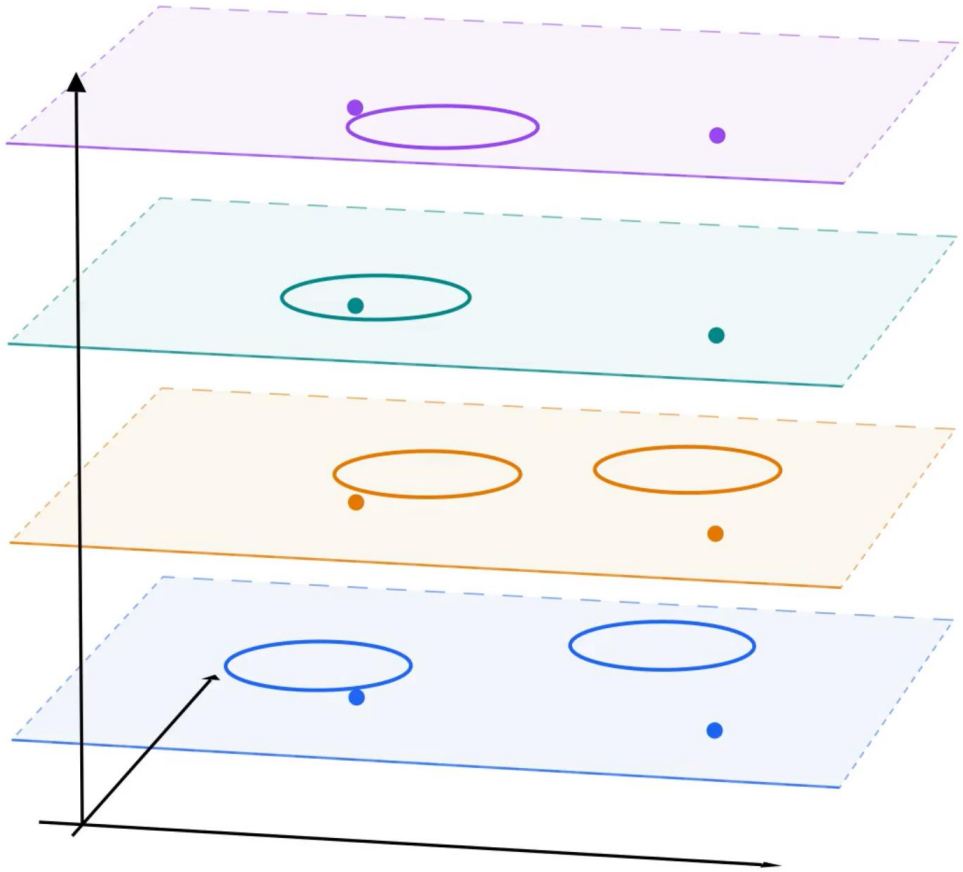




Layer circles by adding an extra dimension



Add colored copies of each point



Concluding Message

- Fuzzy PSI is a cool new technology that can (potentially) solve real-world privacy challenges.
- Not quite there yet though…
- Lots of open questions:
  - A better transformation? Non black-box solutions?
  - FHE-based solutions?
  - ZK Proofs to get malicious security?
  - Parameter choices for applications?
  - MPC + ORAM based solutions?

Thanks!