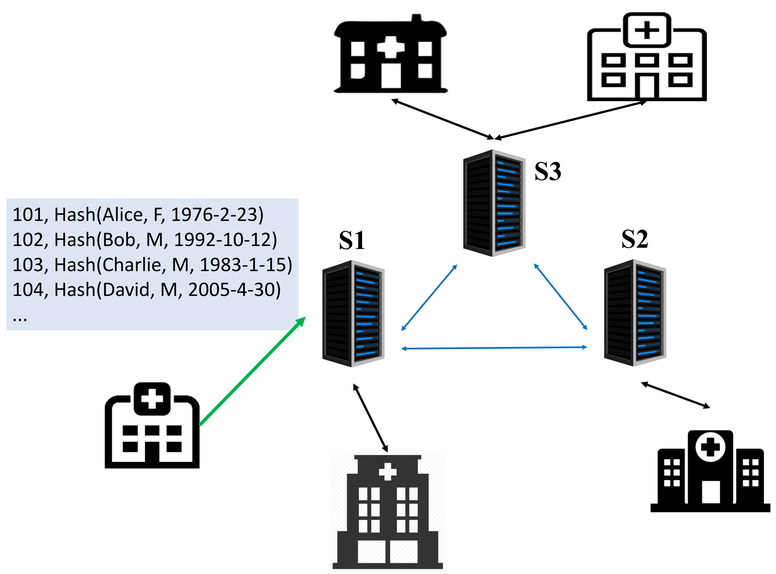
Medical Record De-duplication with APSI

Hao Chen[1], Kyoohyung Han[2], Zhicong Huang[3], Amir Jalali[4], Kim Laine[1], Kristin Lauter[1]

[1]Microsoft Research, [2]Seoul National University, [3]École Polytechnique Fédérale de Lausanne, [4]Florida Atlantic University

Medical Record De-duplication

In this batch of records, which of them exist in other hospitals?



Setting

- 1000 hospitals
- Each hospital has 10, 000 records
- 3~5 servers
- Hospitals send records in batches of 100 to a server
- Sequential and synchronized
- Semi-honest model
- Honest majority among servers
- Hospitals don't collude with servers

Step 1: Asymmetric Private Set Intersection [1]

Server (S1)

Batch X
$\boldsymbol{d}$ records
E.g., $\boldsymbol{d}$ = 100

Cuckoo hashing
3 hash functions
h_1, h_2, h_3

x_1
x_2
x_3
x_4
⋮
x_m

SEAL homomorphic encryption

Cuckoo hash table
table size: $\boldsymbol{m}$
Bin size: **1**

$E(x_1)$
$E(x_2)$
$E(x_3)$
$E(x_4)$
⋮
$E(x_m)$

Server (S2 and S3)

Database Y
$\boldsymbol{D}$ records
$\boldsymbol{D} \gg \boldsymbol{d}$
E.g., $\boldsymbol{D}$=1 million

Simple hashing

Hash each record with all hash functions
$\forall y, h_1(y), h_2(y), h_3(y)$

y_{11}	y_{12}	y_{13}	...	y_{1B}
y_{21}	y_{22}	y_{23}	...	y_{2B}
y_{31}	y_{32}	y_{33}	...	y_{3B}
y_{41}	y_{42}	y_{43}	...	y_{4B}
...	...	...	...	...
y_{m1}	y_{m2}	y_{m3}	...	y_{mB}

Simple hash table
table size: $\boldsymbol{m}$; Bin size: **B**

Server (S1)

$E(x_1)$
$E(x_2)$
$E(x_3)$
$E(x_4)$
⋮
$E(x_m)$

$$E\left(z_j = r_j \prod_{i=1}^{B} (x_j - y_{ji})\right)$$
(Homomorhpic Computation)

$E(z_j)$

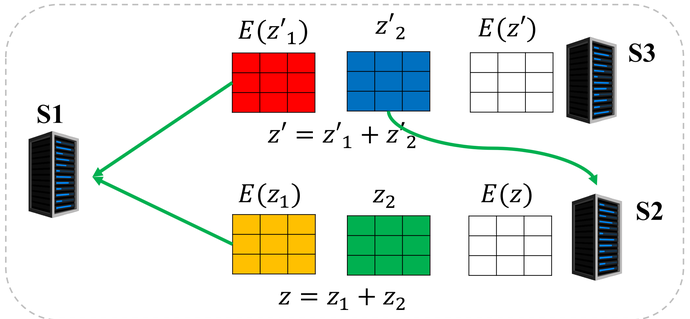
Server (S2 and S3)

y_{11}	y_{12}	y_{13}	...	y_{1B}
y_{21}	y_{22}	y_{23}	...	y_{2B}
y_{31}	y_{32}	y_{33}	...	y_{3B}
y_{41}	y_{42}	y_{43}	...	y_{4B}
...	...	...	...	...
y_{m1}	y_{m2}	y_{m3}	...	y_{mB}

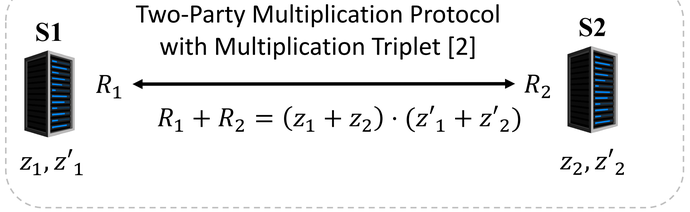
Optimization:
Symmetric polynomial; Splits; Generalized batching; Multithreading

Step 2: Secret-Sharing Set Union

Sharing



Combining



Extending to arbitrary number of servers:
GMW protocol

Results

[1] H. Chen, K. Laine, and P. Rindal, "Fast Private Set Intersection from Homomorphic Encryption", *CCS*, 2017.

[2] D. Demmler, T. Schneider, and M. Zohner, "ABY-A Framework for Efficient Mixed-Protocol Secure Two-Party Computation", NDSS, 2015.