NIST WPEC 2024

Circuit-PSI & Applications

Kyoohyung Han[1], **Seongkwang Kim[1]**,
Byeonghak Lee[1], Yongha Son[2]

[1] Samsung SDS
[2] Sungshin Women's University

Sep. 24. 2024

SAMSUNG SDS

I. Circuit-PSI

Circuit-PSI

Each parties may not want to reveal the intersection itself

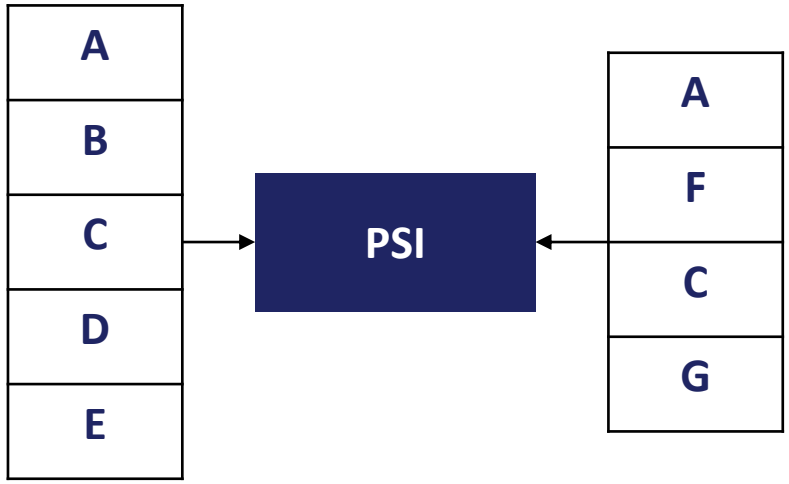
Basic PSI

Circuit-PSI

Circuit-PSI

Each parties may not want to reveal the intersection itself

Basic PSI

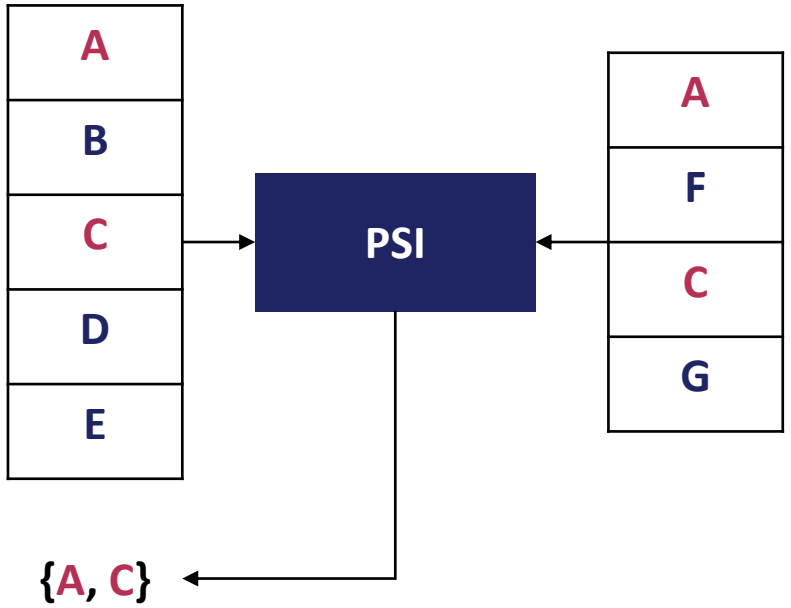


Circuit-PSI

Circuit-PSI

Each parties may not want to reveal the intersection itself

Basic PSI

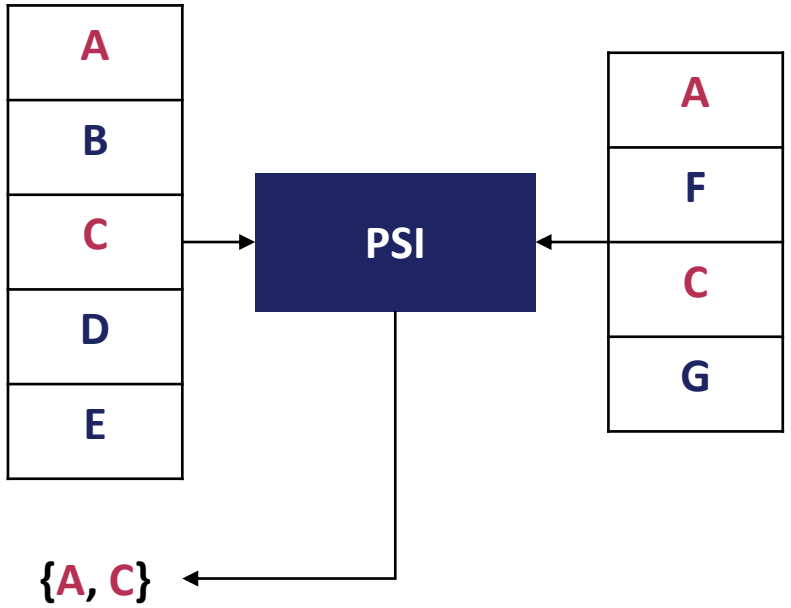


Circuit-PSI

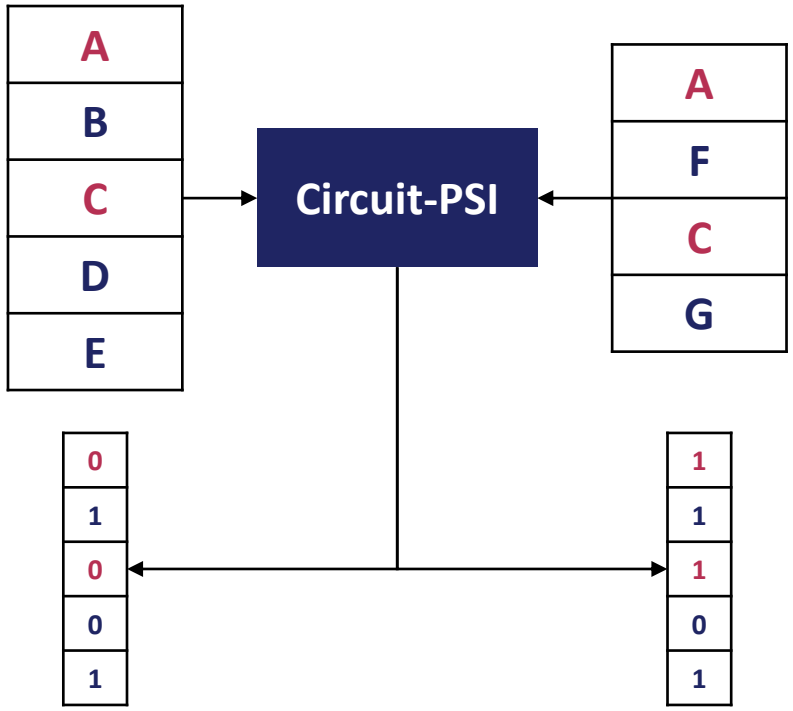
Circuit-PSI

Each parties may not want to reveal the intersection itself

Basic PSI



Circuit-PSI

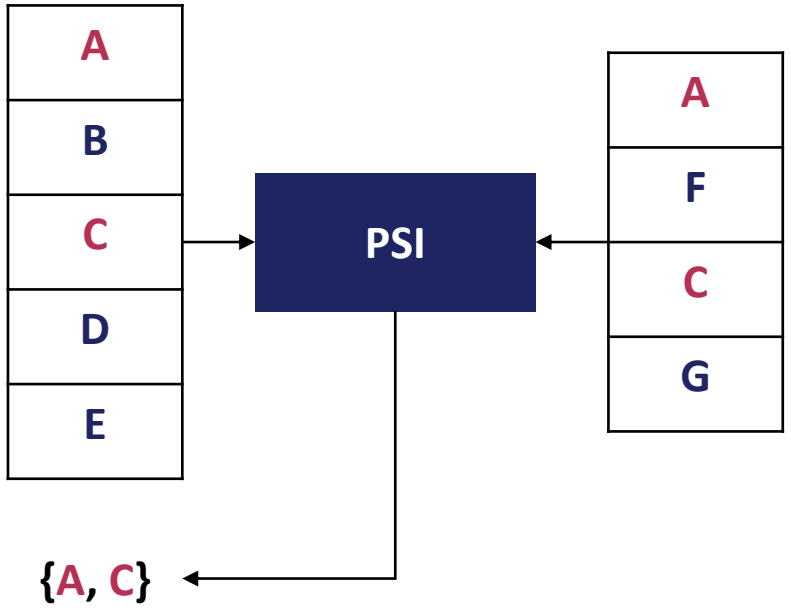


- The XOR sum of two result vector is whether each element is in the other party's set of not

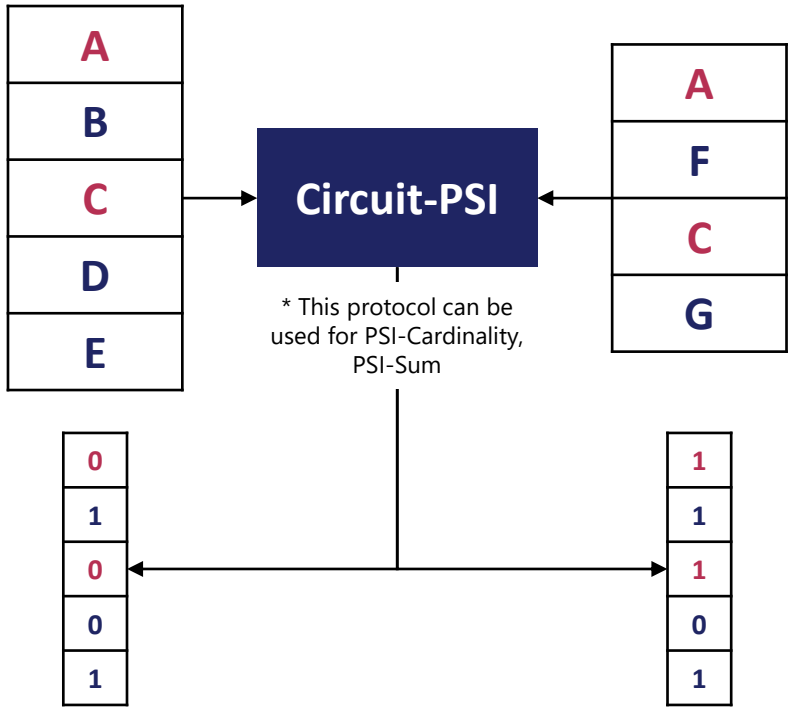
Circuit-PSI

Each parties may not want to reveal the intersection itself

Basic PSI



Circuit-PSI



- The XOR sum of two result vector is whether each element is in the other party's set of not

Brief Mechanism of Circuit-PSI

The circuit-PSI protocol consists of Cuckoo hashing, Oblivious PRF

Brief Mechanism of Circuit-PSI #1

The circuit-PSI protocol consists of Cuckoo hashing, Oblivious PRF

Step1. Cuckoo Hashing

Receiver

A
B
C
D
E

A
F
C
G

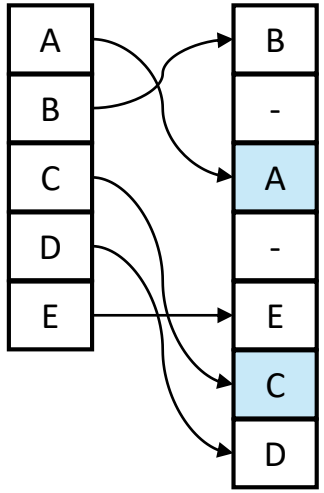
Sender

Brief Mechanism of Circuit-PSI #1

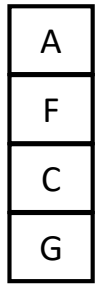
The circuit-PSI protocol consists of Cuckoo hashing, Oblivious PRF

Step1. Cuckoo Hashing

Receiver



Cuckoo Hashing



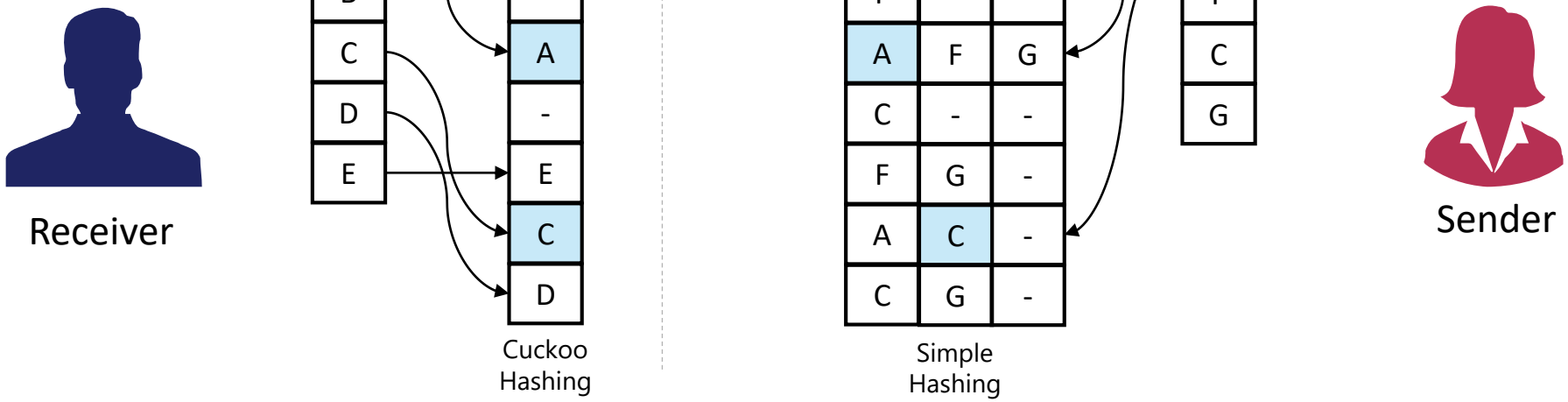
Sender

- The receiver compute Cuckoo hash table, and the sender compute Simple hash table

Brief Mechanism of Circuit-PSI #1

The circuit-PSI protocol consists of Cuckoo hashing, Oblivious PRF

Step1. Cuckoo Hashing

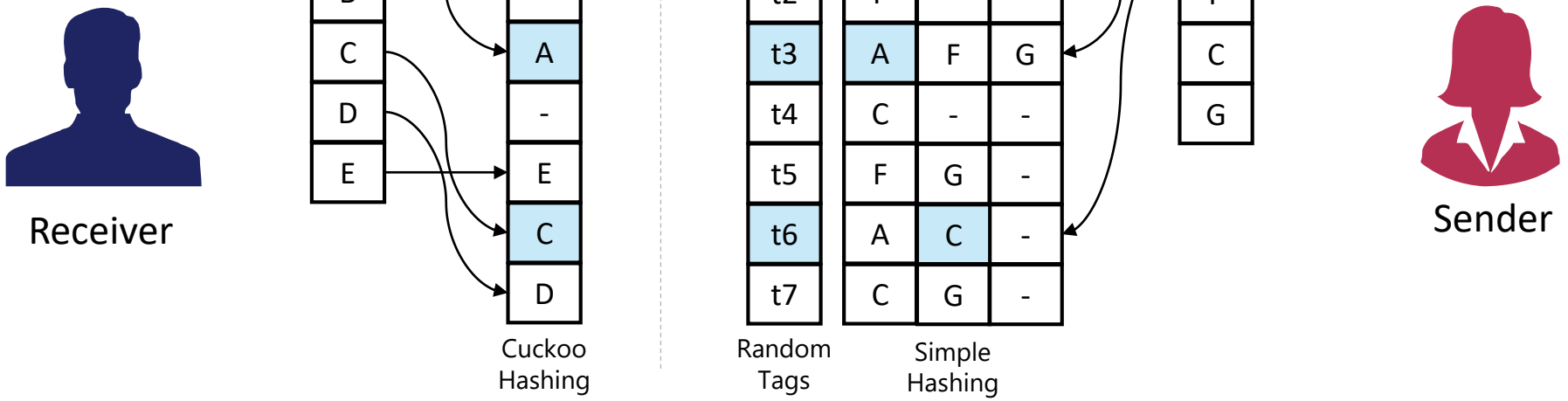


- The receiver compute Cuckoo hash table, and the sender compute Simple hash table

Brief Mechanism of Circuit-PSI #1

The circuit-PSI protocol consists of Cuckoo hashing, Oblivious PRF

Step1. Cuckoo Hashing

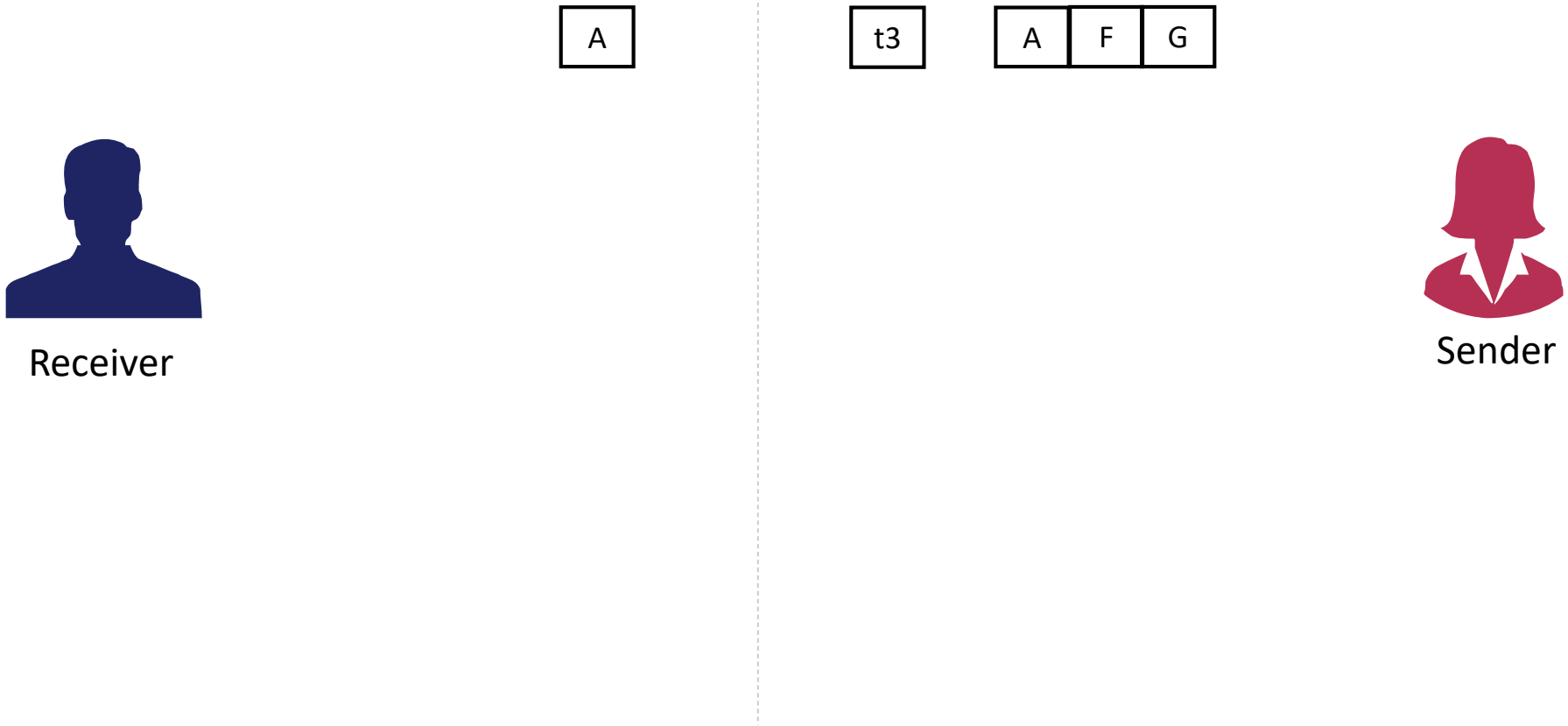


- The receiver compute Cuckoo hash table, and the sender compute Simple hash table
- The sender picks a random tag for each bin of Cuckoo hash table

Brief Mechanism of Circuit-PSI #2

The circuit-PSI protocol consists of Cuckoo hashing, Oblivious PRF

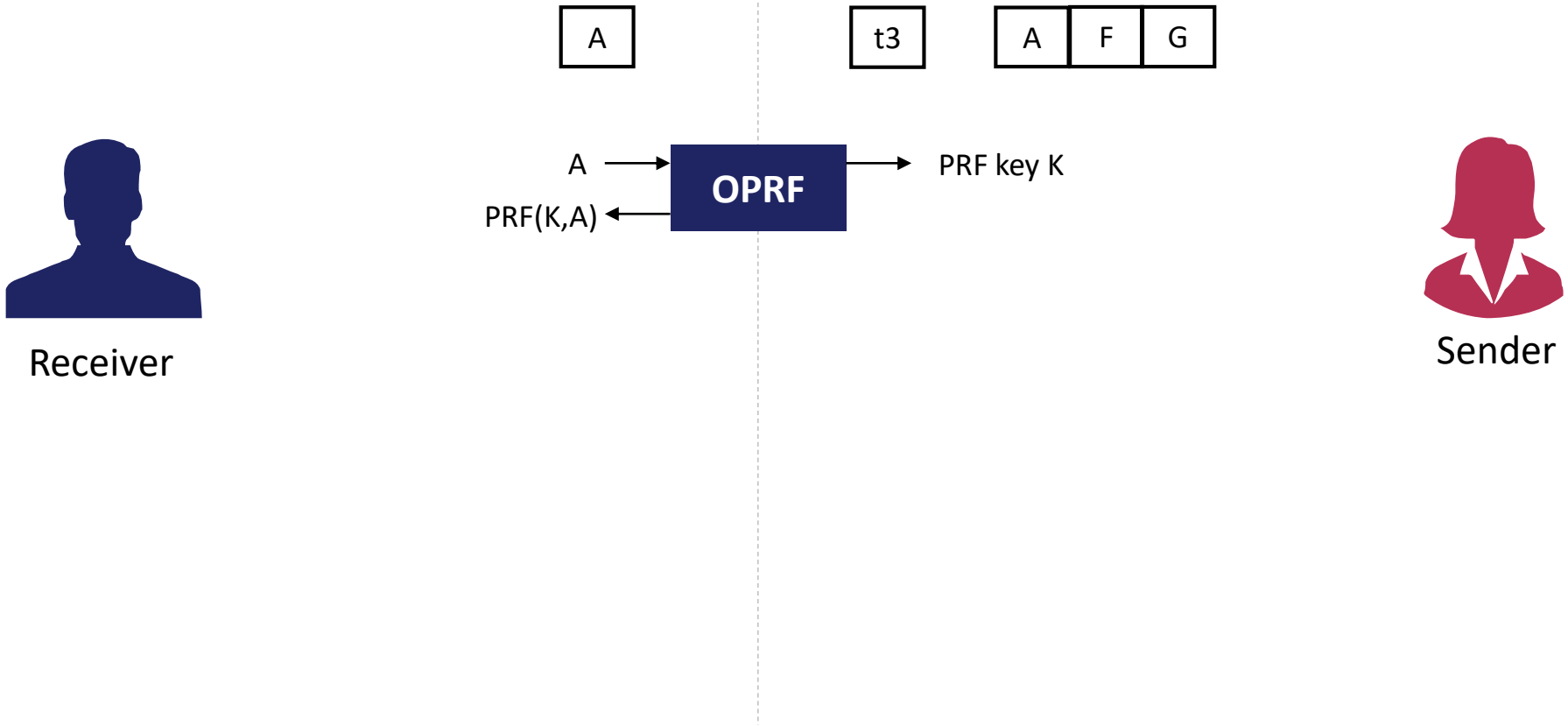
Step2. Oblivious PRF



Brief Mechanism of Circuit-PSI #2

The circuit-PSI protocol consists of Cuckoo hashing, Oblivious PRF

Step2. Oblivious PRF



Brief Mechanism of Circuit-PSI #2

The circuit-PSI protocol consists of Cuckoo hashing, Oblivious PRF

Step2. Oblivious PRF

Receiver

A

A →

← PRF(K,A)

OPRF

→ PRF key K

t3

A	F	G

Sender

Compute and send r1, …, r4:

r1 = PRF(K,F) + t3	H(F) mod 4 = 0
r2 = PRF(K,G) + t3	H(G) mod 4 = 1
r3 = PRF(K,A) + t3	H(A) mod 4 = 2
r4 = random	

Brief Mechanism of Circuit-PSI #2

The circuit-PSI protocol consists of Cuckoo hashing, Oblivious PRF

Step2. Oblivious PRF

Receiver

A

A →

← PRF(K,A)

OPRF

→ PRF key K

t3

A	F	G

Sender

Compute and send r1, …, r4:

r1 = PRF(K,F) + t3	H(F) mod 4 = 0
r2 = PRF(K,G) + t3	H(G) mod 4 = 1
r3 = PRF(K,A) + t3	H(A) mod 4 = 2
r4 = random	

① Compute H(A) mod 4 = 3

② Compute s3 = OPRF(A) + r3 = t3

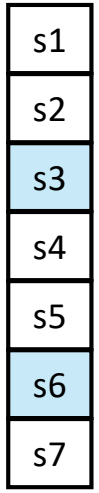
: s3 and t3 is same if and only if "A" is in other party's set

Brief Mechanism of Circuit-PSI #3

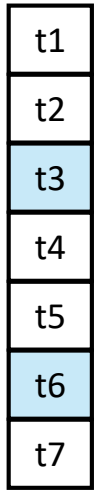
The circuit-PSI protocol consists of Cuckoo hashing, Oblivious PRF

Step3. Private Equality Test

Receiver



Computed Tags



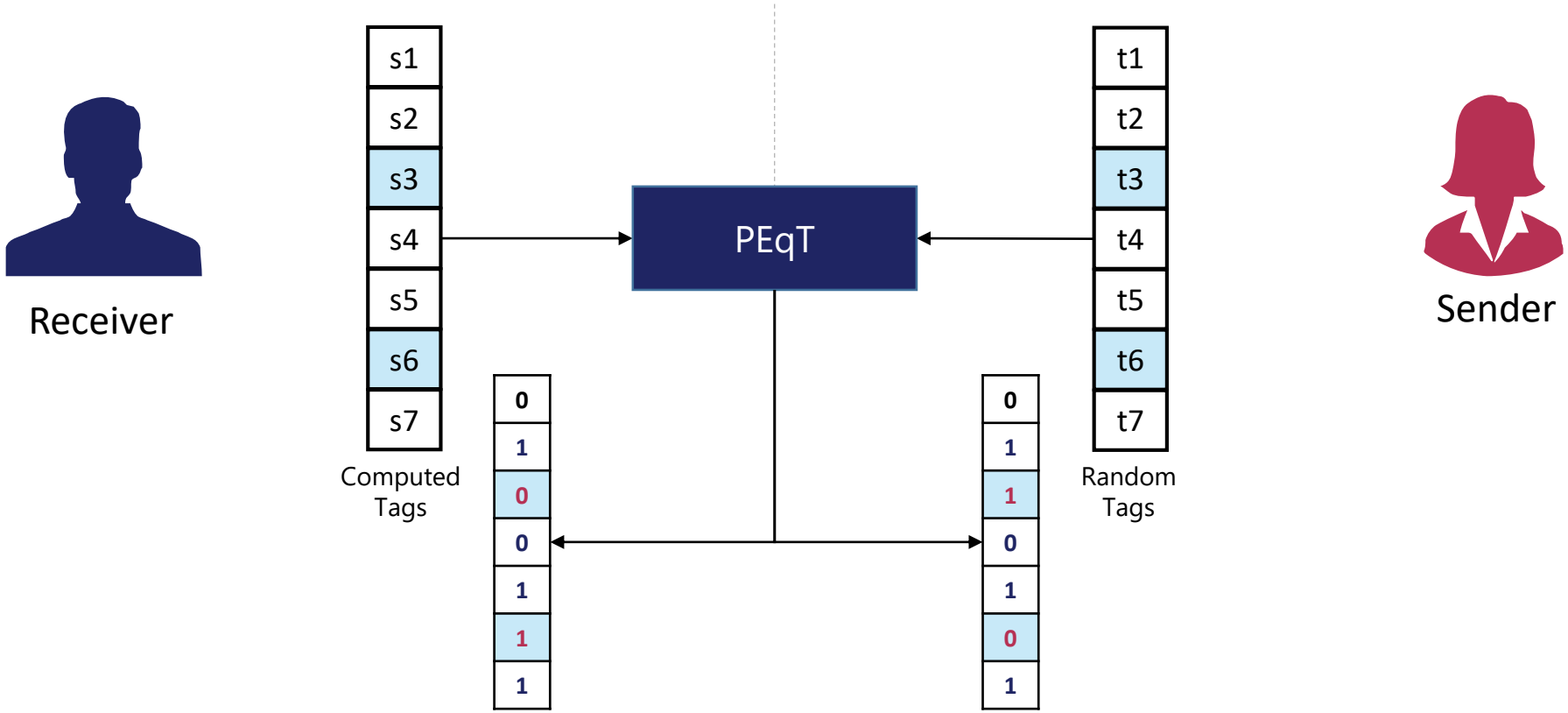
Random Tags

Sender

Brief Mechanism of Circuit-PSI #3

The circuit-PSI protocol consists of Cuckoo hashing, Oblivious PRF

Step3. Private Equality Test



Our Works on Circuit-PSI

HE-based optimizations, and security of building block

Our Works on Circuit-PSI

HE-based optimizations, and security of building block

CPSI with HE (SAC '22)

- **Reduce communication cost during Step3: equality preserving compression**
  - Compute below equation using HE scheme
  - For $x = \sum x_i B^i, y = \sum y_i B^i$,

$$r + \sum (x_i - y_i)^2 = r \text{ iff } x = y$$

* Ferret-OT is used for OT extension

Network	# of Items	Comm. (MB)	Time (s)
LAN	2^{16}	20.69	2.09
	2^{18}	83.64	5.71
	2^{20}	394.3	20.17

Table 1. Performance of Circuit-PSI

[HMS22] K. Han, D. Moon, and Y. Son. *Improved Circuit-PSI via Equality Preserving Compression*. SAC 2022.

Our Works on Circuit-PSI

HE-based optimizations, and security of building block

CPSI with HE (SAC '22)

- **Reduce communication cost during Step3: equality preserving compression**
  - Compute below equation using HE scheme
  - For $x = \sum x_i B^i, y = \sum y_i B^i$,

$$r + \sum (x_i - y_i)^2 = r \text{ iff } x = y$$

* Ferret-OT is used for OT extension

Network	# of Items	Comm. (MB)	Time (s)
LAN	2^{16}	20.69	2.09
	2^{18}	83.64	5.71
	2^{20}	394.3	20.17

Table 1. Performance of Circuit-PSI

[HMS22] K. Han, D. Moon, and Y. Son. *Improved Circuit-PSI via Equality Preserving Compression*. SAC 2022.

Unbalanced CPSI with HE (AsiaCCS '23)

- **Convert HE-based PSI to CPSI**
  - Zero storage required for small set holder
  - Recursive HE application for trade-off

[SJ23] Y. Son, and J. Jeong. *PSI with computation or Circuit-PSI for Unbalanced Sets from Homomorphic Encryption*. AsiaCCS 2023.

Our Works on Circuit-PSI

HE-based optimizations, and security of building block

CPSI with HE (SAC '22)

- **Reduce communication cost during Step3: equality preserving compression**
  - Compute below equation using HE scheme
  - For $x = \sum x_i B^i, y = \sum y_i B^i$,

$$r + \sum (x_i - y_i)^2 = r \text{ iff } x = y$$

* Ferret-OT is used for OT extension

Network	# of Items	Comm. (MB)	Time (s)
LAN	2^{16}	20.69	2.09
	2^{18}	83.64	5.71
	2^{20}	394.3	20.17

Table 1. Performance of Circuit-PSI

[HMS22] K. Han, D. Moon, and Y. Son. *Improved Circuit-PSI via Equality Preserving Compression*. SAC 2022.

Unbalanced CPSI with HE (AsiaCCS '23)

- **Convert HE-based PSI to CPSI**
  - Zero storage required for small set holder
  - Recursive HE application for trade-off

For Better Building Block (AC '24)

- **Cryptanalysis on Malicious OPRF**
  - OPRF for PSI is generally OTe (or VOLE) + OKVS
  - We can get at most double OPRF evaluations by overfitting OKVS
  - It leads PSI to statistical distance larger than $2^{-\lambda}$

[SJ23] Y. Son, and J. Jeong. *PSI with computation or Circuit-PSI for Unbalanced Sets from Homomorphic Encryption*. AsiaCCS 2023.
[HKLS24] K. Han, S. Kim, B. Lee, and Y. Son. *Revisiting OKVS-based OPRF and PSI: Cryptanalysis and Better Construction*. Asiacrypt 2024, to appear.

II. Applications

1. Genomic Analysis

Safe Sharing of Genomic Data

Pathogen genomic data is generally considered to be non-identifying, but it can be combined with personal information

Scenario

Results

Safe Sharing of Genomic Data

Pathogen genomic data is generally considered to be non-identifying, but it can be combined with personal information

Scenario

Has Tet A	ID
Yes	Abc02
No	Def01
No	Abc01
Yes	Def02

ID	Homeless
Abc01	Yes
Abc02	Yes
Abc03	Yes
Abc04	No

Results

Safe Sharing of Genomic Data

Pathogen genomic data is generally considered to be non-identifying, but it can be combined with personal information

Scenario

Has Tet A	ID
Yes	Abc02
No	Def01
No	Abc01
Yes	Def02

Remain
Has Tet A = Yes

ID
Abc02
Def02

ID	Homeless
Abc01	Yes
Abc02	Yes
Abc03	Yes
Abc04	No

Remain
Homeless = Yes

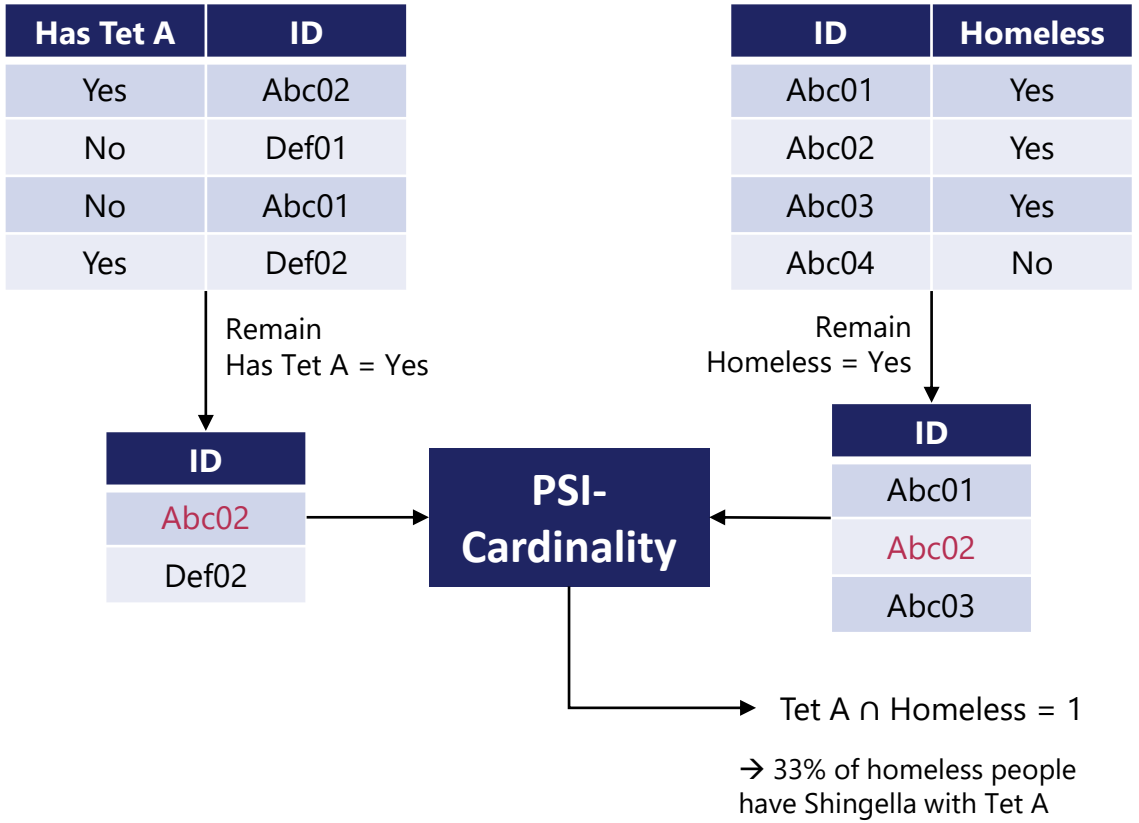
ID
Abc01
Abc02
Abc03

Results

Safe Sharing of Genomic Data

Pathogen genomic data is generally considered to be non-identifying, but it can be combined with personal information

Scenario

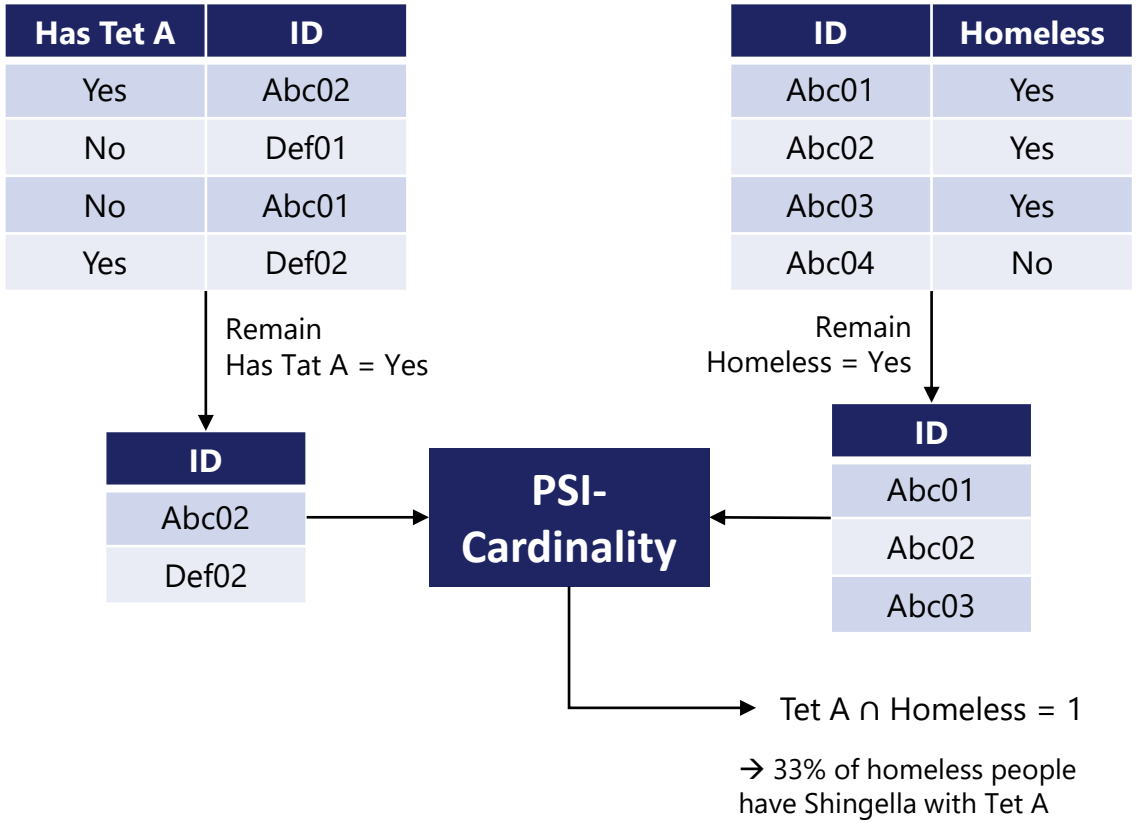


Results

Safe Sharing of Genomic Data

Pathogen genomic data is generally considered to be non-identifying, but it can be combined with personal information

Scenario



Results

Outputs

- 76.92% of homeless people and 0% of non-homeless people have Shingella bacteria with Tet A

Timing Result

- For 30 rows, it only takes 1 second.
- We estimate that it takes less than a minute for 1M rows

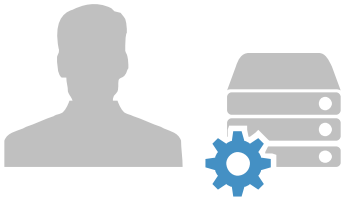
Safe Sharing of Genomic Data from Discovery to Broad Data Sharing Leveraging the Power of Beacon and Privacy Enhancing Technology (PET), Galaxy Community Conference 2023 (w/ Soyean Kim from SFU)

2. Data Aggregation

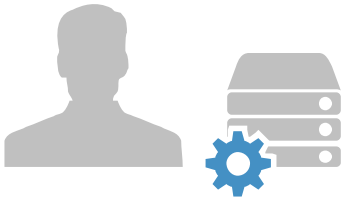
Data Aggregation System in Korea

Highly sensitive identifier only goes to trusted 3rd party (government agency)

TTP



Party A



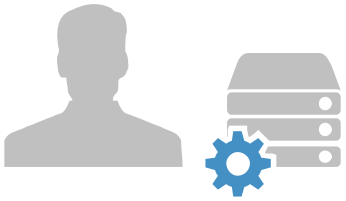
Party B

Data aggregation center
(DAC)

Data Aggregation System in Korea

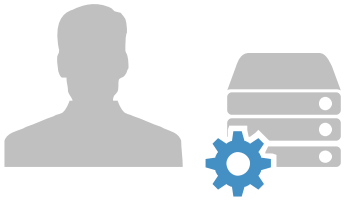
Highly sensitive identifier only goes to trusted 3rd party (government agency)

TTP



Party A
Data includes:
IDs, indices, features

Index	ID (phone)	ZIP	DOB	...
A01	010-1234-5678	12345	1978-SEP-21	...
A02	010-2345-6789	56789	1996-JUN-05	...



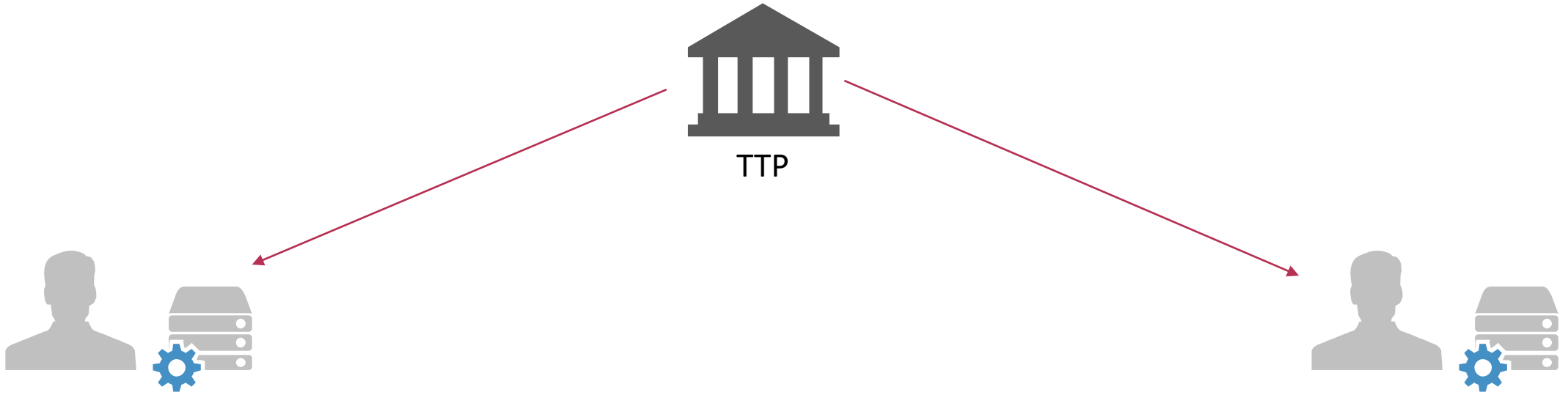
Party B
Data includes:
IDs, indices, features

Data aggregation center
(DAC)

Data Aggregation System in Korea

1. TTP sends salt to parties

Highly sensitive identifier only goes to trusted 3rd party (government agency)



Party A
Data includes:
IDs, indices, features

Index	ID (phone)	ZIP	DOB	...
A01	010-1234-5678	12345	1978-SEP-21	...
A02	010-2345-6789	56789	1996-JUN-05	...

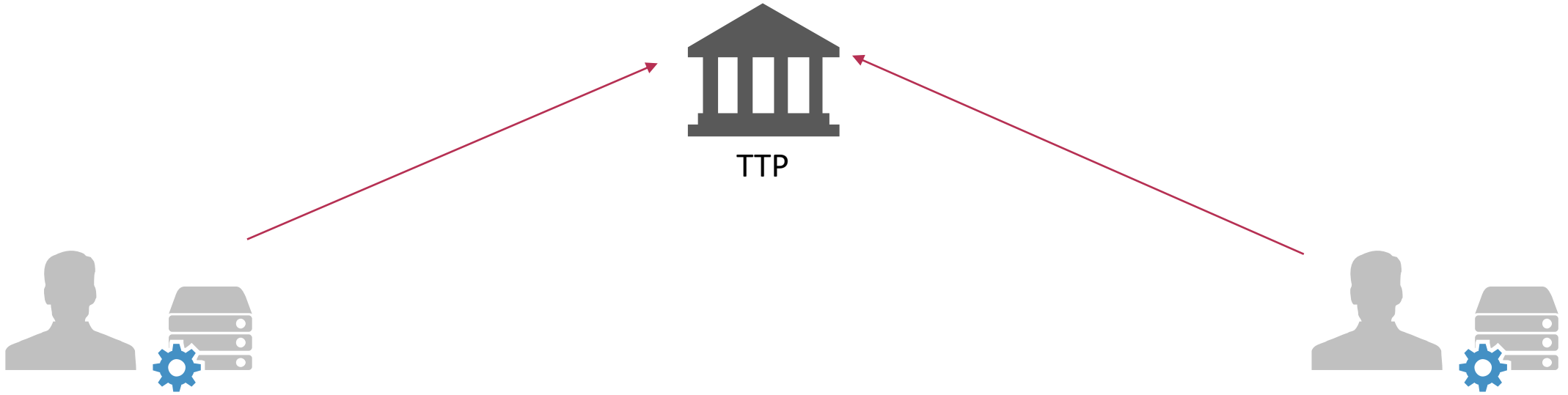
Party B
Data includes:
IDs, indices, features

Data aggregation center (DAC)

Data Aggregation System in Korea

Highly sensitive identifier only goes to trusted 3rd party (government agency)

1. TTP sends salt to parties
2. Parties send H(ID), and indices to TTP



Party A
Data includes:
IDs, indices, features

Index	ID (phone)	ZIP	DOB	...
A01	010-1234-5678	12345	1978-SEP-21	...
A02	010-2345-6789	56789	1996-JUN-05	...

Party B
Data includes:
IDs, indices, features

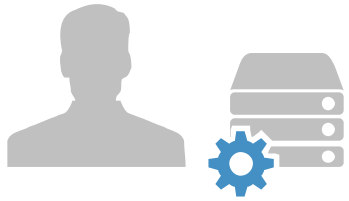
Data aggregation center (DAC)

Data Aggregation System in Korea

Highly sensitive identifier only goes to trusted 3rd party (government agency)

1. TTP sends salt to parties
2. Parties send H(ID), and indices to TTP
3. TTP sends index map to DAC

TTP

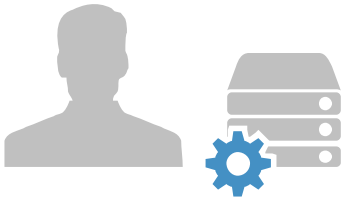


Party A
Data includes:
IDs, indices, features

Index	ID (phone)	ZIP	DOB	...
A01	010-1234-5678	12345	1978-SEP-21	...
A02	010-2345-6789	56789	1996-JUN-05	...

Index A	Index B
A01	B85
A02	B07
...	...

Data aggregation center (DAC)



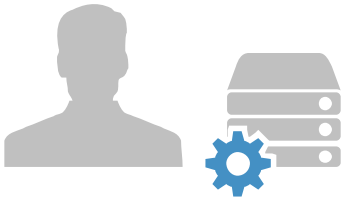
Party B
Data includes:
IDs, indices, features

Data Aggregation System in Korea

Highly sensitive identifier only goes to trusted 3rd party (government agency)

1. TTP sends salt to parties
2. Parties send H(ID), and indices to TTP
3. TTP sends index map to DAC
4. DAC sends matching ratio to the parties

TTP



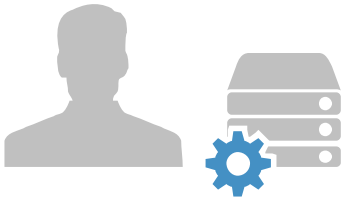
Party A
Data includes:
IDs, indices, features

Index	ID (phone)	ZIP	DOB	...
A01	010-1234-5678	12345	1978-SEP-21	...
A02	010-2345-6789	56789	1996-JUN-05	...

Matching ratio = 45%

Data aggregation center (DAC)

Matching ratio = 45%



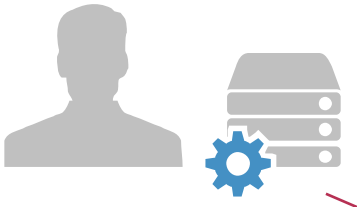
Party B
Data includes:
IDs, indices, features

Data Aggregation System in Korea

Highly sensitive identifier only goes to trusted 3rd party (government agency)

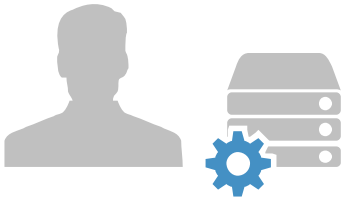
TTP

1. TTP sends salt to parties
2. Parties send H(ID), and indices to TTP
3. TTP sends index map to DAC
4. DAC sends matching ratio to the parties
5. Parties send indices, and features to DAC



Party A
Data includes:
IDs, indices, features

Index	ID (phone)	ZIP	DOB	...
A01	010-1234-5678	12345	1978-SEP-21	...
A02	010-2345-6789	56789	1996-JUN-05	...



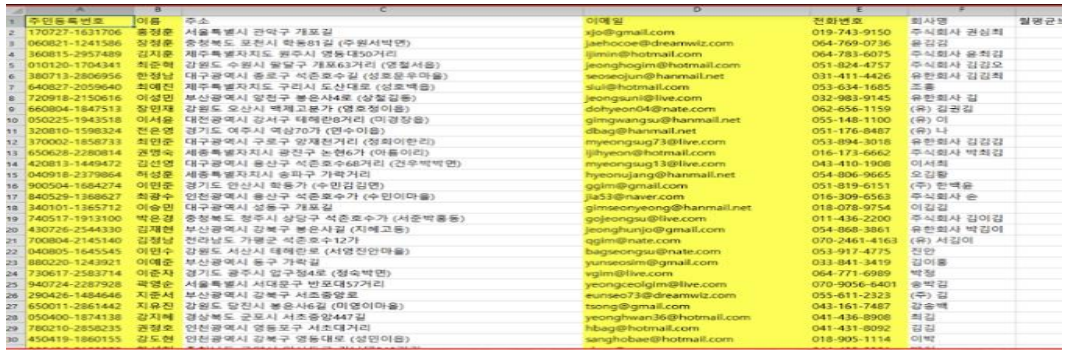
Party B
Data includes:
IDs, indices, features

Data aggregation center (DAC)

Circuit-PSI based Data Aggregation System

Samsung SDS is one of the data aggregation center designated by government

Example of Outputs



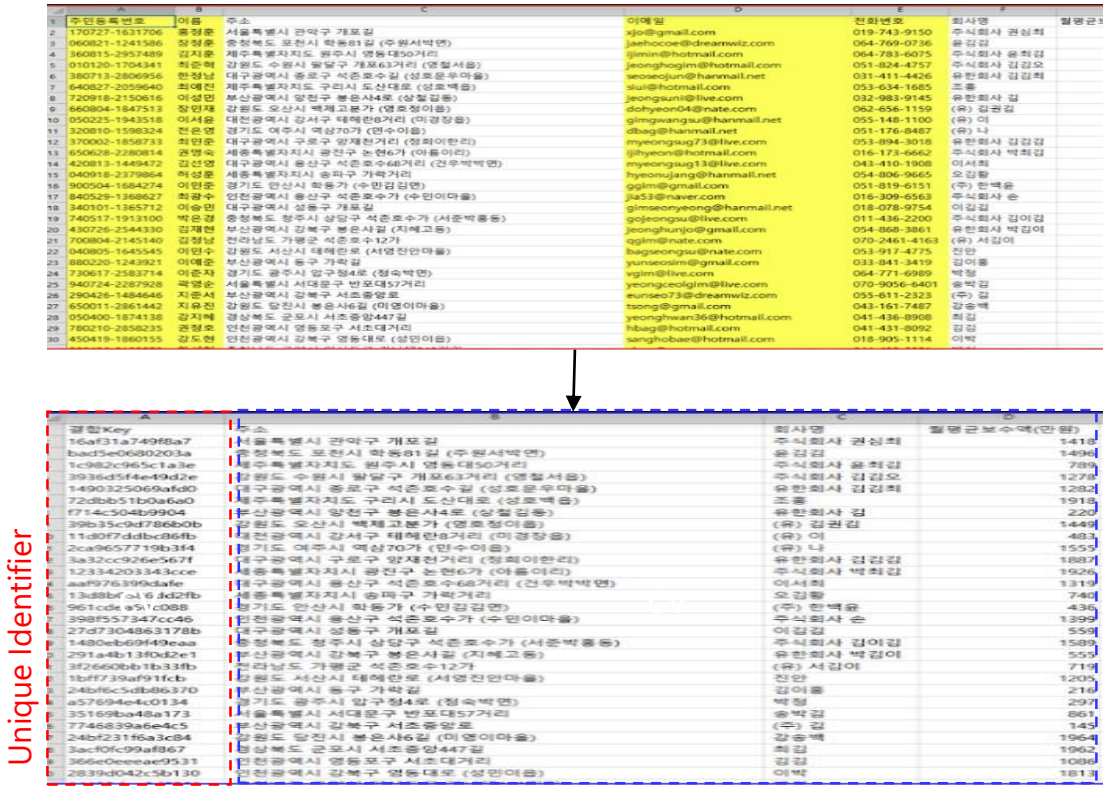
	A	B	C	D	E	F
1	주민등록번호	이름	주소	이메일	전화번호	회사명
2	170727-1631706	홍정훈	서울특별시 관악구 개포길	xjo@gmail.com	019-743-9150	주식회사 권심희
3	060821-1241586	장정훈	충청북도 포천시 학동81길 (주원서박면)	jaehocoe@dreamwiz.com	064-769-0736	윤김김
4	360815-2957489	김지훈	제주특별자치도 원주시 영동대50거리	ijimin@hotmail.com	064-783-6075	주식회사 윤최김
5	010120-1704341	최준혁	강원도 수원시 일달구 개포63거리 (연철서윤)	jeonghogim@hotmail.com	051-824-4757	주식회사 김김오
6	380713-2806956	한정남	대구광역시 종로구 석촌호수길 (성호윤우마을)	seoseojun@hanmail.net	031-411-4426	유한회사 김김최
7	640827-2059640	최예진	제주특별자치도 구리시 도산대로 (성호백읍)	siuj@hotmail.com	053-634-1685	조홍
8	720918-2150616	이성민	부산광역시 양천구 봉은사4로 (상철김동)	jeongsuni@live.com	032-983-9145	유한회사 김
9	660804-1847513	장연재	강원도 오산시 백제고분가 (영호정이읍)	dohyeon04@nate.com	062-656-1159	(유) 김권김
10	050225-1943518	이서윤	대전광역시 강서구 테헤란8거리 (미경장읍)	gimgwangsu@hanmail.net	055-148-1100	(유) 이
11	320810-1598324	전은영	경기도 여주시 역삼7071 (민수이읍)	dbag@hanmail.net	051-176-8487	(유) 나
12	370002-1858733	최연준	대구광역시 구로구 압재천거리 (정희이한리)	myeongsug73@live.com	053-894-3018	유한회사 김김김
13	650628-2280814	권명숙	세종특별자치시 광진구 논현6가 (아름이리)	ijihyeon@hotmail.com	016-173-6662	주식회사 박최김
14	420813-1449472	김선영	대구광역시 용산구 석촌호수68거리 (건우박박면)	myeongsug13@live.com	043-410-1908	이서희
15	040918-2379864	허성훈	세종특별자치시 송파구 가락거리	hyeonujang@hanmail.net	054-806-9665	오김황
16	900504-1684274	이연준	경기도 안산시 학동가 (수민김읍)	ggjm@gmail.com	051-819-6151	(주) 한백윤
17	840529-1368627	최광수	인천광역시 용산구 석촌호수가 (수민이마을)	jia53@naver.com	016-309-6563	주식회사 손
18	340101-1365712	이승민	대구광역시 성동구 개포길	gimseonyeong@hanmail.net	018-078-9754	이김김
19	740517-1913100	박은경	충청북도 청주시 상당구 석촌호수가 (서준박종동)	gojeongsu@live.com	011-436-2200	주식회사 김이김
20	430726-2544330	김재현	부산광역시 강북구 봉은사길 (지혜고동)	jeonghunjo@gmail.com	054-868-3861	유한회사 박김이
21	700804-2145140	김정남	전라남도 가평군 석촌호수12가	qgjm@nate.com	070-2461-4163	(유) 서김이
22	040805-1645545	이연수	강원도 서산시 태헤란로 (서영잔안마을)	bagseongsu@nate.com	053-917-4775	잔안
23	880220-1243921	이예준	부산광역시 동구 가락길	yunseoslm@gmail.com	033-841-3419	김이홍
24	730617-2583714	이준차	경기도 광주시 압구정4로 (정숙박면)	ygjm@live.com	064-771-6989	박정
25	940724-2287928	곽영순	서울특별시 서대문구 반포대57거리	yeongceolgim@live.com	070-9056-6401	송박김
26	290426-1484646	지준서	부산광역시 강북구 서초중앙로	eunseo73@dreamwiz.com	055-611-2323	(주) 김
27	650011-2861442	지유진	강원도 당진시 봉은사6길 (미연이마을)	tsong@gmail.com	043-161-7487	강송백
28	050400-1874138	강지혁	경상북도 군포시 서초중앙447길	yeonghwan36@hotmail.com	041-436-8908	최김
29	780210-2858235	권정호	인천광역시 영동로구 서초대거리	hbag@hotmail.com	041-431-8092	김김
30	450419-1860155	강도현	인천광역시 강북구 영동대로 (성민이읍)	sanghobae@hotmail.com	018-905-1114	이박

Performances

Circuit-PSI based Data Aggregation System

Samsung SDS is one of the data aggregation center designated by government

Example of Outputs



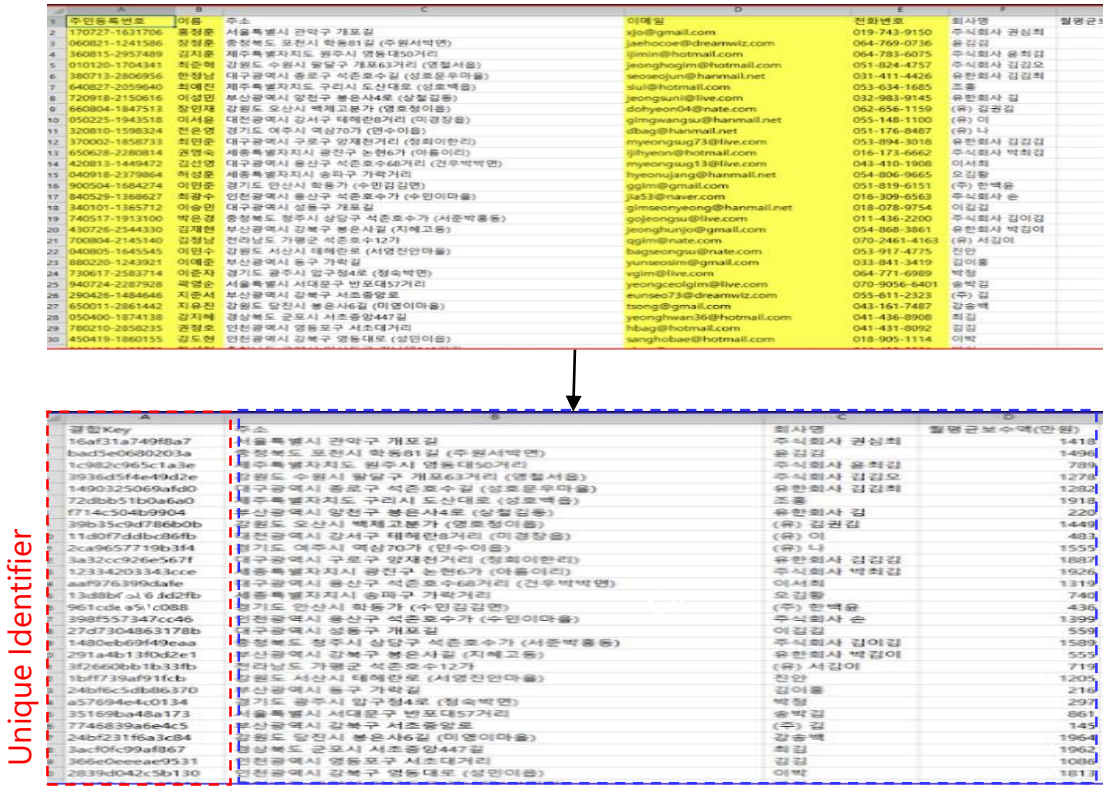
Matching ratio : 0.43 (43% of your data is in other party's data)

Performances

Circuit-PSI based Data Aggregation System

Samsung SDS is one of the data aggregation center designated by government

Example of Outputs



Matching ratio : 0.43 (43% of your data is in other party's data)

Performances

How it works

- For the unique ID generation, we used OPRF instead of salted hash
- Matching ratio is computed using circuit-PSI

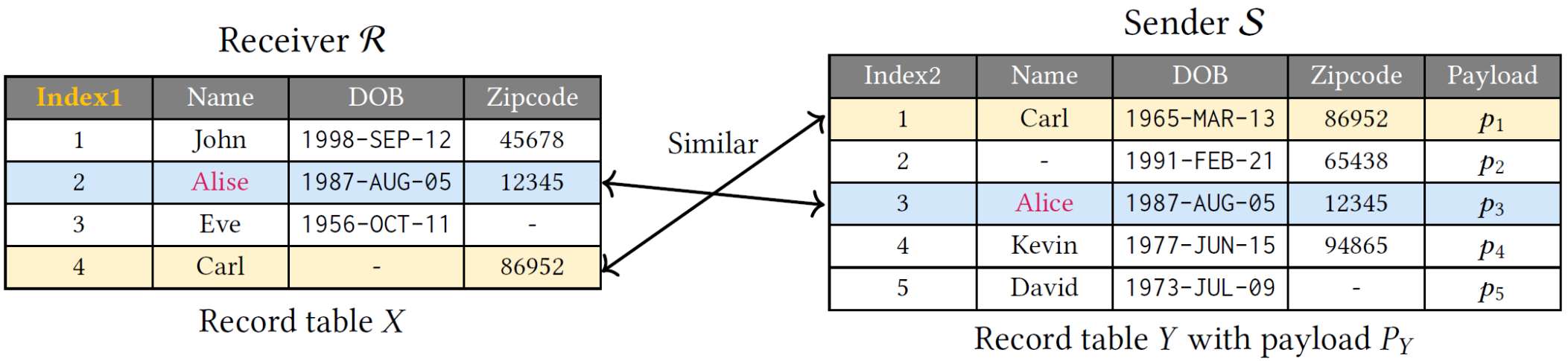
Timing Result (LAN)

	Dataset Size	Running Time
Join Key Generation & Ratio Computation	1m	1m 28s
	10m	11m 38s
	20m	20m 40s
Multiple ID Col Support	-	Checked

3. Privacy-Preserving Record Linkage

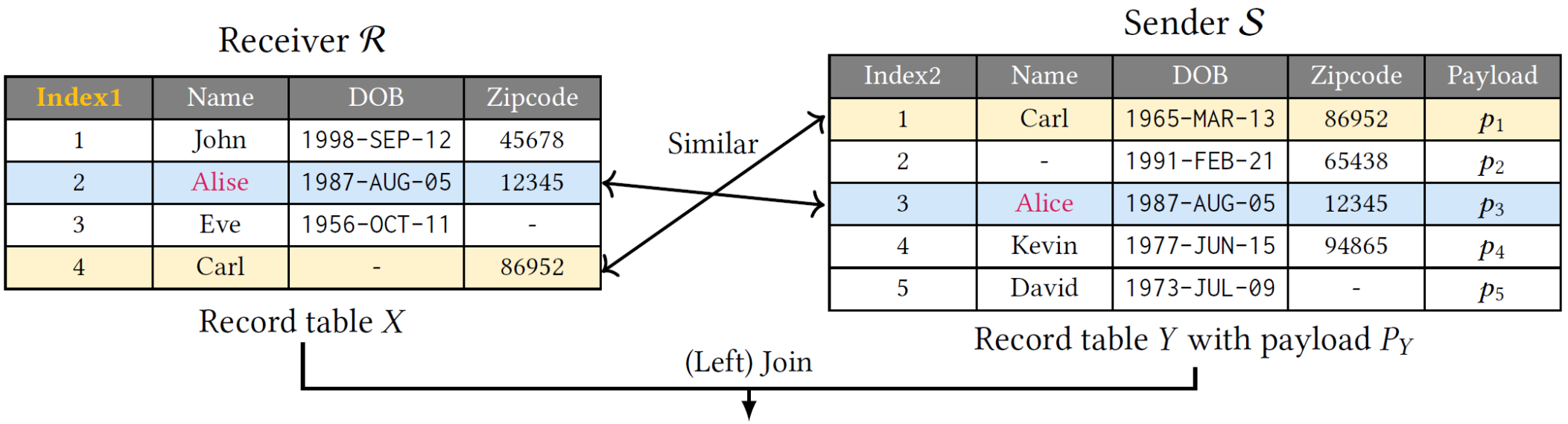
Privacy-Preserving Record Linkage (PETS '25)

What is target functionality?



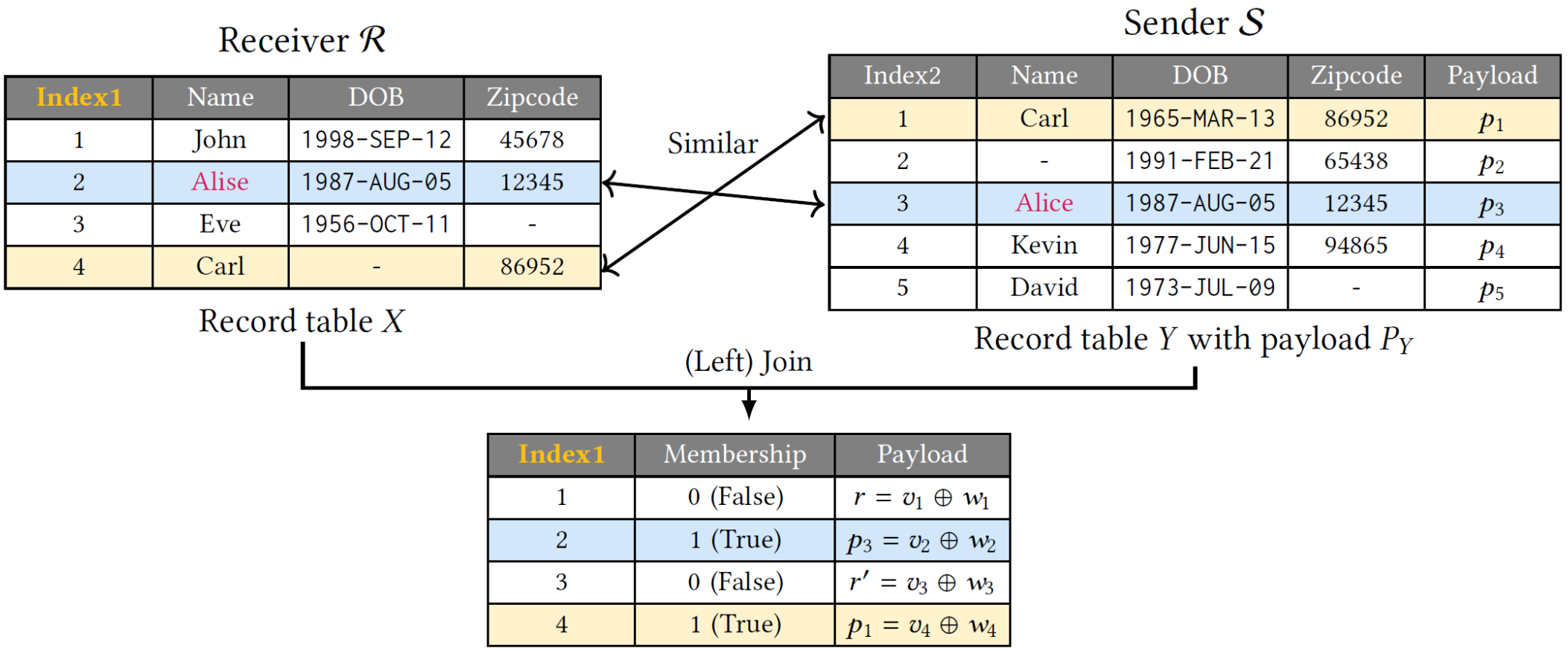
Privacy-Preserving Record Linkage (PETS '25)

What is target functionality?



Privacy-Preserving Record Linkage (PETS '25)

What is target functionality?



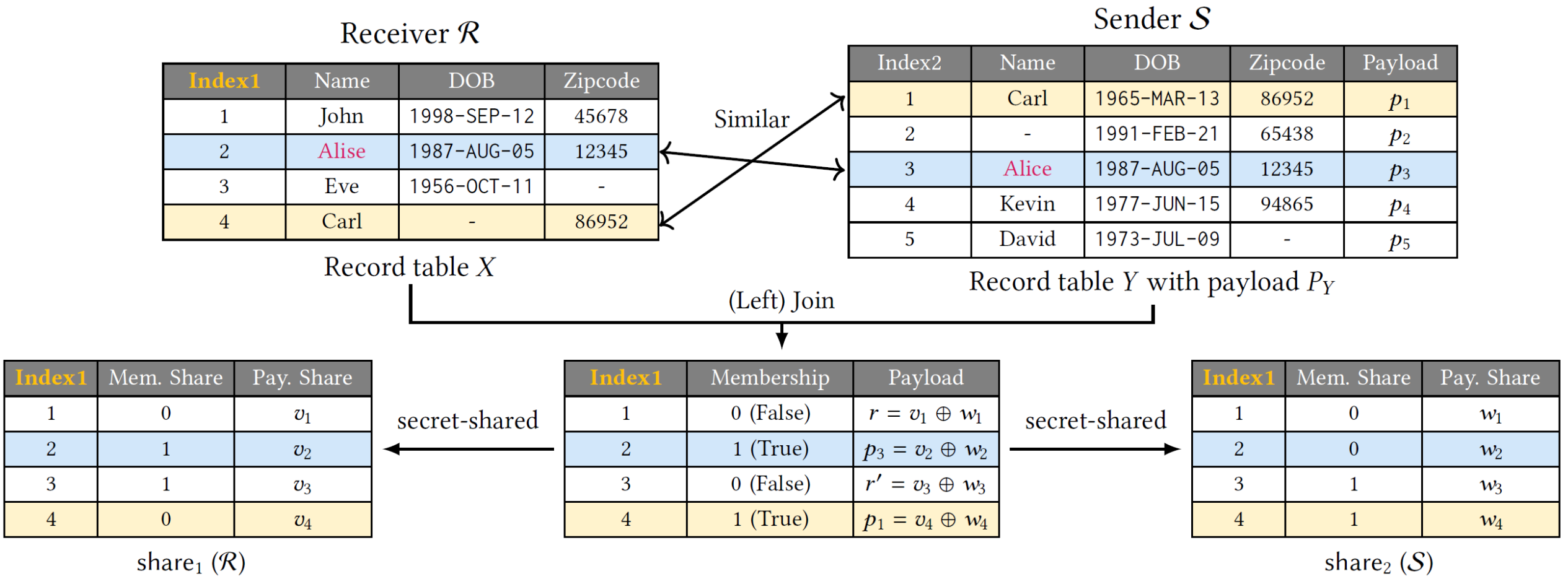
Index1	Name	DOB	Zipcode
1	John	1998-SEP-12	45678
2	Alise	1987-AUG-05	12345
3	Eve	1956-OCT-11	-
4	Carl	-	86952

Index2	Name	DOB	Zipcode	Payload
1	Carl	1965-MAR-13	86952	p_1
2	-	1991-FEB-21	65438	p_2
3	Alice	1987-AUG-05	12345	p_3
4	Kevin	1977-JUN-15	94865	p_4
5	David	1973-JUL-09	-	p_5

Index1	Membership	Payload
1	0 (False)	$r = v_1 \oplus w_1$
2	1 (True)	$p_3 = v_2 \oplus w_2$
3	0 (False)	$r' = v_3 \oplus w_3$
4	1 (True)	$p_1 = v_4 \oplus w_4$

Privacy-Preserving Record Linkage (PETS '25)

What is target functionality?



Privacy-Preserving Record Linkage (PETS '25)

from CPSI and some MPC techniques (simplified)

1. Encode features (e.g., concatenation, LSH)

Index1	Name	DOB	Zipcode
1	John	1998-SEP-12	45678
2	Alise	1987-AUG-05	12345
3	Eve	1956-OCT-11	-
4	Carl	-	86952

Record table X with quasi-identifiers

Encode

Index1	feature1	feature2	feature3	feature4
1	d330ca	719cf0	0e0348	79457c
2	01191e	bf9a47	0c1452	c41b44
3	e60467	b91fc3	0e79d1	005dae
4	f6fabd	a8fdd6	a80394	737963

Encoded table $\hat{X}$ with features

Privacy-Preserving Record Linkage (PETS '25)

from CPSI and some MPC techniques (simplified)

1. Encode features (e.g., concatenation, LSH)
2. Invoke CPSI protocols feature-wisely

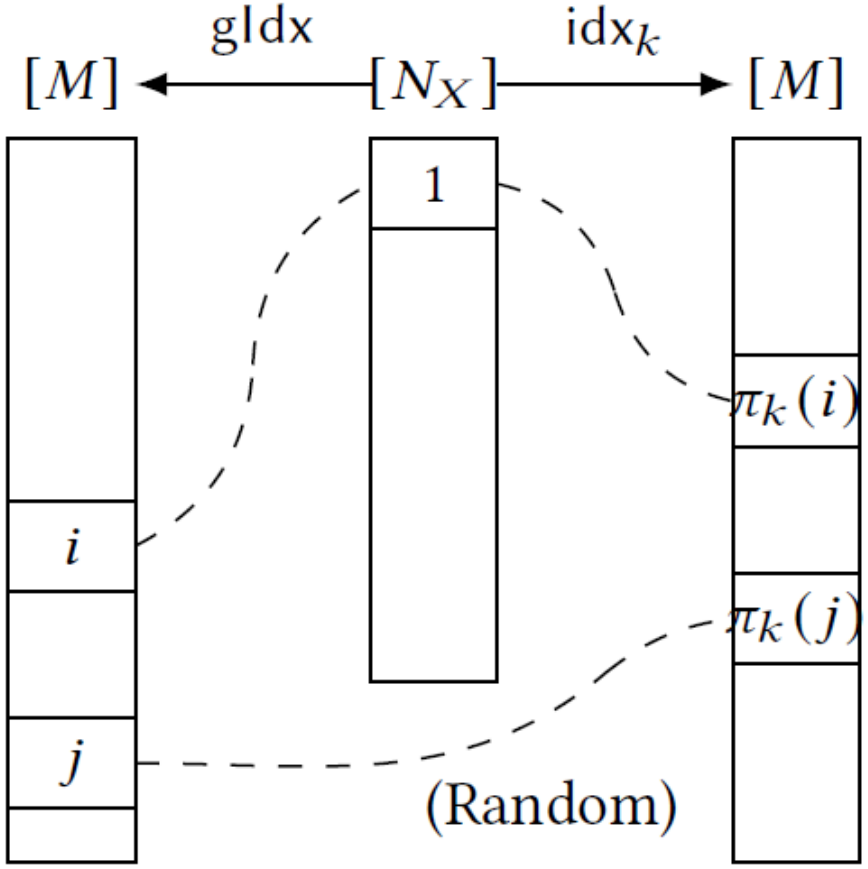
	CPSI	CPSI	CPSI	CPSI
Index1	feature1	feature2	feature3	feature4
1	d330ca	719cf0	0e0348	79457c
2	01191e	bf9a47	0c1452	c41b44
3	e60467	b91fc3	0e79d1	005dae
4	f6fabd	a8fdd6	a80394	737963

Encoded table $\hat{X}$ with features

Privacy-Preserving Record Linkage (PETS '25)

from CPSI and some MPC techniques (simplified)

1. Encode features (e.g., concatenation, LSH)
2. Invoke CPSI protocols feature-wisely
3. Align the shares using Permute-and-Share protocol



Privacy-Preserving Record Linkage (PETS '25)

from CPSI and some MPC techniques (simplified)

1. Encode features (e.g., concatenation, LSH)
2. Invoke CPSI protocols feature-wisely
3. Align the shares using Permute-and-Share protocol
4. Compute matching criterion using generic MPC (e.g., GMW)

$$\sum_i \mathrm{Eq}(\mathrm{feature}_i) \geq t$$

Index1	feature1	feature2	feature3	feature4
1	d330ca	719cf0	0e0348	79457c
2	01191e	bf9a47	0c1452	c41b44
3	e60467	b91fc3	0e79d1	005dae
4	f6fabd	a8fdd6	a80394	737963

Encoded table $\hat{X}$ with features

Privacy-Preserving Record Linkage (PETS '25)

from CPSI and some MPC techniques (simplified)

1. Encode features (e.g., concatenation, LSH)
2. Invoke CPSI protocols feature-wisely
3. Align the shares using Permute-and-Share protocol
4. Compute matching criterion using generic MPC (e.g., GMW)

$$\sum_i \mathrm{Eq}(\mathrm{feature}_i) \geq t$$

Dataset	Data Size	#Features	Comm. (MB)	Setup (s)	Online (s)	F1-score
European Census	24K	6	76.2	4.47	1.13	0.948
NCVR	1M	3	1615	169	23.7	0.976

References

- Data table in p.37-39 is generated by Faker (owner: Francois Zaninotto, url: https://faker.readthedocs.io/en/master/) at Oct. 1. 2021
- [HMS22] K. Han, D. Moon, and Y. Son. Improved Circuit-PSI via Equality Preserving Compression. SAC 2022.
- [SJ23] Y. Son, and J. Jeong. PSI with computation or Circuit-PSI for Unbalanced Sets from Homomorphic Encryption. AsiaCCS 2023.
- [HKLS24] K. Han, S. Kim, B. Lee, and Y. Son. Revisiting OKVS-based OPRF and PSI: Cryptanalysis and Better Construction. Asiacrypt 2024, to appear.
- [HKS24] K. Han, S. Kim, and Y. Son. Private Computation on Common Fuzzy Records. PETS 2025, to appear.

Thank you

Q&A

SAMSUNG SDS