UNECE

modernstats by HLG - MOS

Input Privacy-Preservation Techniques Project

Privacy Set Intersection with Analytics

***Massimo De Cubellis**, Mauro Bruno, Fabrizio De Fausti, Monica Scannapieco (ISTAT)*

Definitions

Private Set Intersection (PSI) is a problem within the field of Secure multi-party computation.

Secure multi-party computation (SMPC) is a subfield of cryptography with the goal of creating methods for parties to jointly compute a function over their inputs while keeping those inputs private.

The PSI problem:

There are two friends Alice and Bob such that Alice has a set of items A=(a1,…,an) and Bob has the set B=(b1,…,bn).

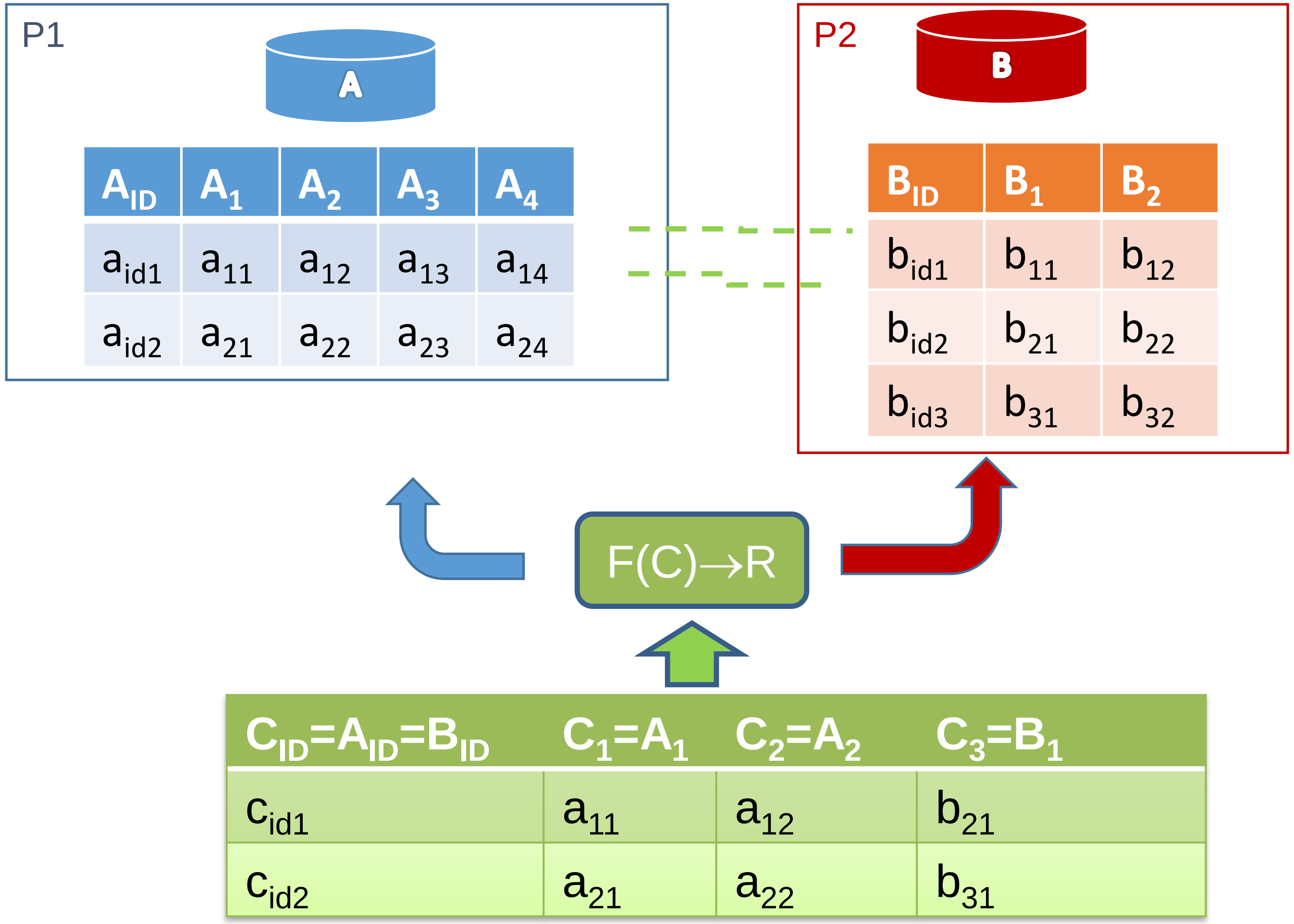
The goal is to design a protocol by which Alice and Bob obtain the intersection A∩B, under the restriction that the protocol must not reveal anything about items that are not in the intersection.

Scenarios for Privacy Set Intersection

Four scenarios:

- Private Set Intersection (PSI)
- Private Set Intersection with Enrichment (PSI-E)
- **Private Set Intersection with Analytics (PSI-A)**
- Private data mining (PDM)

PSI – Private Set Intersection with Analytics



A_{ID}	A_1	A_2	A_3	A_4
a_{id1}	a_{11}	a_{12}	a_{13}	a_{14}
a_{id2}	a_{21}	a_{22}	a_{23}	a_{24}

B_{ID}	B_1	B_2
b_{id1}	b_{11}	b_{12}
b_{id2}	b_{21}	b_{22}
b_{id3}	b_{31}	b_{32}

$C_{ID}=A_{ID}=B_{ID}$	$C_1=A_1$	$C_2=A_2$	$C_3=B_1$
c_{id1}	a_{11}	a_{12}	b_{21}
c_{id2}	a_{21}	a_{22}	b_{31}

$C\ (C_{ID}=A_{ID}=B_{ID},\ C_1=A_1, C_2=A_2,\ C_3=A_3)=A \cap B=\{c_{id1}=a_{id1}=b_{id2}, c_{id2}=a_{id2}=b_{id3}\}$

Private Set Intersection with Analytics Case Study: Istat and Bank of Italy

- Two phases of the protocol: offline and online
- Offline phase:
    - The two parties want to share datasets of individuals that have a common key that is the Fiscal Code.
    - Agree to share some variables:
        - Istat: *Number of Children, Age Class*
        - Bank of Italy: *Income class, type of mortgage payment, payer type*
    - Share a *symmetric key* through RSA protocol
    - Share IP addresses to use

Case Study: Istat and Bank of Italy

- Online phase:
  - Private Set Intersection
    - E. De Cristofaro and G Tsudik
      Practical Private Set Intersection Protocols with linear Computational and Bandwidth Complexity.
      proc Financial Cryptography and data Security, 2010.
  - Loading: encrypted data transmission to the Linker;
  - Query: submission of queries to the Linker and transmission of results.

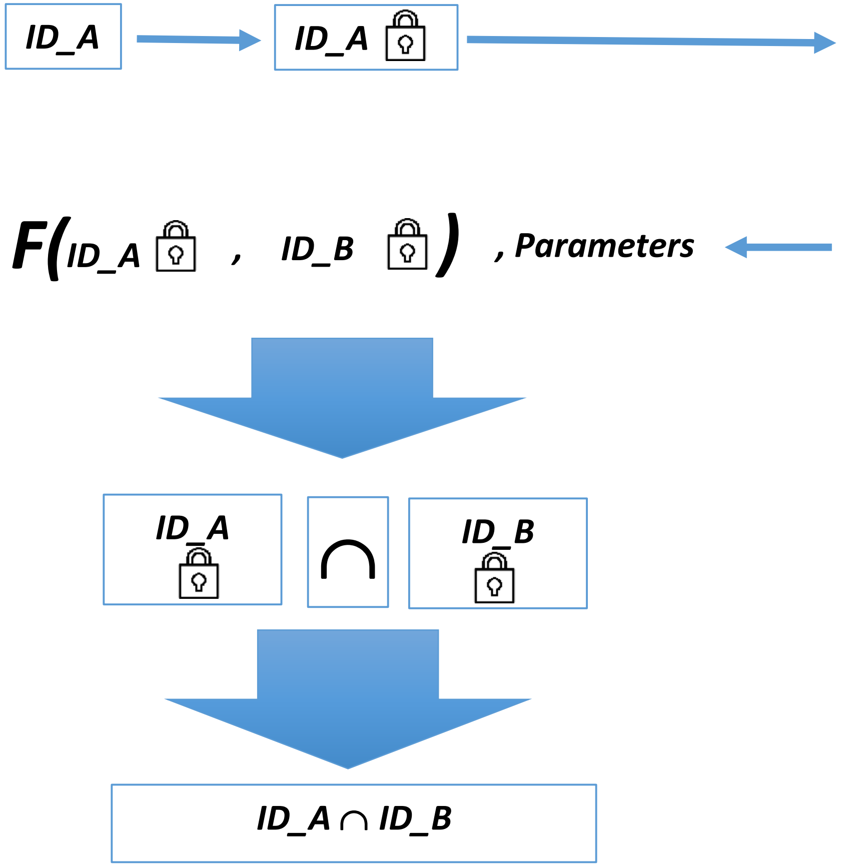
Private Set Intersection phase

ISTAT (Client)

(Datataset A: ID_A, A_1, A_2)

BANK of ITALY (Server)

(Datataset B: ID_B, B_1, B_2, B_3)



ID_A → *ID_A* 🔒 →

F(ID_A 🔒 ***, ID_B*** 🔒 ***)*** ***, Parameters*** ←

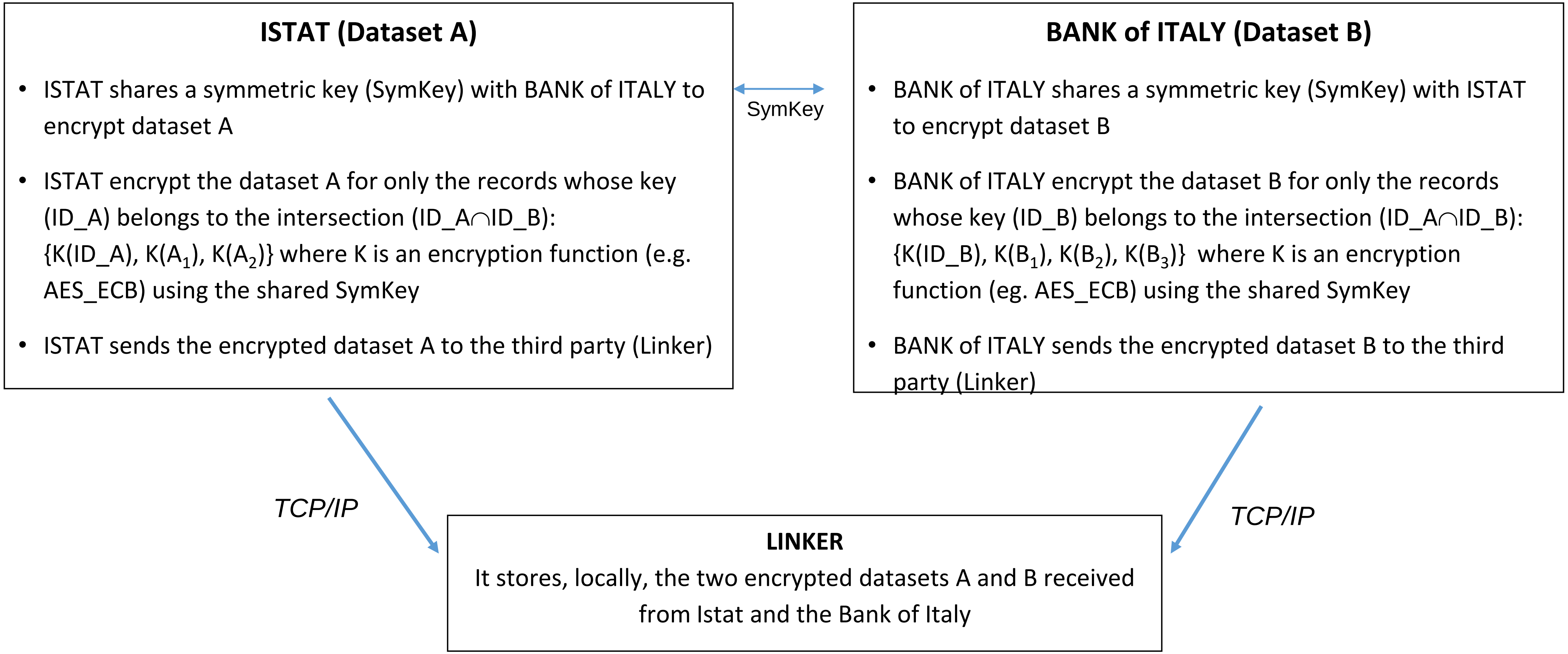
ID_A 🔒 ∩ *ID_B* 🔒

ID_A ∩ ID_B

- Calculate a function that use as input: ID_A encrypted, ID_B encrypted, and a set of parameters
- Send the function to ISTAT

Loading phase

The encrypted datasets of the two parts are uploaded on the Linker server



ISTAT (Dataset A)

- ISTAT shares a symmetric key (SymKey) with BANK of ITALY to encrypt dataset A
- ISTAT encrypt the dataset A for only the records whose key (ID_A) belongs to the intersection (ID_A∩ID_B): {K(ID_A), K(A_1), K(A_2)} where K is an encryption function (e.g. AES_ECB) using the shared SymKey
- ISTAT sends the encrypted dataset A to the third party (Linker)

BANK of ITALY (Dataset B)

- BANK of ITALY shares a symmetric key (SymKey) with ISTAT to encrypt dataset B
- BANK of ITALY encrypt the dataset B for only the records whose key (ID_B) belongs to the intersection (ID_A∩ID_B): {K(ID_B), K(B_1), K(B_2), K(B_3)} where K is an encryption function (eg. AES_ECB) using the shared SymKey
- BANK of ITALY sends the encrypted dataset B to the third party (Linker)

LINKER

It stores, locally, the two encrypted datasets A and B received from Istat and the Bank of Italy

Query phase (analytics)

ISTAT

- ISTAT, in asynchronous way, sends a query to the Linker (eg: queryString = NUMBER_OF_CHILDREN, INCOME_CLASS);
- ISTAT receives the result of the query from the Linker

BANK of ITALY

- BANK of ITALY, in asynchronous way, sends a query to the Linker (eg: queryString = INCOME_CLASS, KIND_OF_MORTGAGE_PAYMENT, AGE_CLASS);
- BANK of ITALY receives the result of the query from the Linker;

TCP/IP
TCP/IP

LINKER

- Receives requests (query group by / counts)
- To perform the required count, it join the IDs of the two encrypted datasets received in the loading phase
- Send the result of the query to the requester

Characteristics

- Transmission to the third party of only the encrypted and strictly necessary data;
- The two parties do not exchange data directly, except those necessary to calculate the intersection on the A∩B keys, but use a neutral third party;
- Information enrichment takes place only in terms of aggregated data (counts);
- The third party can carry out checks on the counts returned and ensure that the result cannot be traced back to the individual elements of the population.
- Privacy preservation is not guaranteed in the event that one of the two parties agrees dishonestly with the third party

References

1) *Emiliano De Cristofaro and Gene Tsudik, Practical Private Set Intersection Protocols with Linear Computational and Bandwidth Complexity, Proceedings of Financial Cryptography and Data Security, 2010*
2) *David Evans, Vladimir Kolesnikov and Mike Rosulek, A Pragmatic Introduction to Secure Multi-Party Computation. NOW Publishers, 2018.*
3) *A. ACAR, H. AKSU, and A. S. LUAGAC, M. CONTI A Survey on Homomorphic Encryption Schemes: Theory and Implementation, 2017,* https://arxiv.org/abs/1704.03578
4) *Sophia Yakoubov ; Vijay Gadepally ; Nabil Schear ; Emily Shen ; Arkady Yerukhimovich, A survey of cryptographic approaches to securing big-data analytics in the cloud IEEE High Performance Extreme Computing Conference (HPEC), 2014*

Thanks for your attention !!!