CoE DSC

KNIGHTS ANALYTICS

Roseman Labs
PRIVACY BY DESIGN

Data collaboration to fight financial crime

Using MPC for entity resolution: a case study

Management summary

1. Financial crime and associated activities cause tremendous harm from both a societal and an economic perspective. Examples are laundering money associated with drug trafficking, forced prostitution and terrorism financing (non-exhaustive)
2. Financial institutions struggle to fulfil their role in the combat against financial crime, for which they have to detect suspicious transactions. This is partly because relevant data is fragmented across multiple sources and banks, making it hard to have a complete view on clients and their transactions
3. Creating a combined view on clients and their transactions by collaborating on data can help financial institutions in their combat against financial crime because it provides them with more complete insights on actual transaction patterns and organisations
4. To create this combined view, records relating to different organisations need to be linked to the same real-world entity. This linking is called entity resolution. Entity resolution can be conducted by either deploying relatively simple rule-based algorithms, or more sophisticated AI algorithms
5. Entity resolution at an interbank level is often not possible because banks aren't allowed to share data containing sensitive or confidential information beyond the boundaries of their own organisation due to privacy regulations, confidentiality agreements and professional secrecy obligations
6. Knights Analytics have developed a solution that utilises AI and can be combined with Multi-Party Computation (MPC) to conduct entity resolution at an interbank level and create a combined view whilst minimising sensitive and/or confidential information being shared. This solution aims to improve the detection of financial crime without compromising sensitive information
7. The Centre of Excellence – Data Sharing and Cloud (CoE-DSC) has supported Knights Analytics to assess how the introduction of MPC influences the performance of entity resolution in terms of quality and scalability and for which other use cases the developed solution could be utilised
8. This solution helps in combatting financial crime by creating insights that couldn't be created before. The results show that combining entity resolution with MPC reduces data exposure, while the performance is similar to the solution without MPC in terms of:
   - Quality: When adding pre-resolution as an extra step in the process, the number of entities that have been resolved with MPC is similar to the number of resolved entities when entity resolution is done in the clear (i.e. without privacy boundaries), although some over-resolution is observed
   - Scalability: The solution with MPC requires more computation power than the solution without MPC due to extra steps added by the cryptographic process. This isn't a major problem in the context of detection of financial crime because this increase in computational time is linear, meaning that using MPC won't lead to an insurmountable increase of computational power as the number of records grows
9. The combination of entity resolution with MPC can be beneficial for a range of use cases beyond the scope of financial crime. To illustrate this, examples of applications for child protection, prevention of tax evasion, and detection of insurance fraud are included in the report

Most important terms used in this document

Term	Explanation
AML/CFT regulation	EU legislative framework for anti-money laundering and countering the financing of terrorism (read more here)
Entity resolution	An analysis conducted to match the identity of some real-world entity (e.g. organisation, person) to the corresponding records within and/or across datasets
Interbank entity resolution	An analysis conducted to match records to the real-world entity across datasets of several banks (i.e. analysis is run across multiple banks)
Intrabank entity resolution	An analysis conducted to match records to the real-world entity across datasets of one bank (i.e. analysis is run within a bank)
Multi-Party Computation (MPC)	A type of privacy enhancing technology where computations are securely run at each party ensuring that the source data remains private and only insights are shared
Privacy boundary	A barrier between privacy domains of each organisation involved in computations (e.g. as part of the privacy domain of a bank, there are boundaries that ensure sensitive source data is safeguarded)
Privacy Enhancing Technology (PET)	A technical implementation that enables analysis of data in a way that sensitive data remains protected and secure
Trusted Third Party	A party that is entrusted by data providers to conduct analyses (e.g. in case of AML and CFT monitoring, authorities like ACPR, SAMLIT, TMNL act as Trusted Third Parties in their respective countries

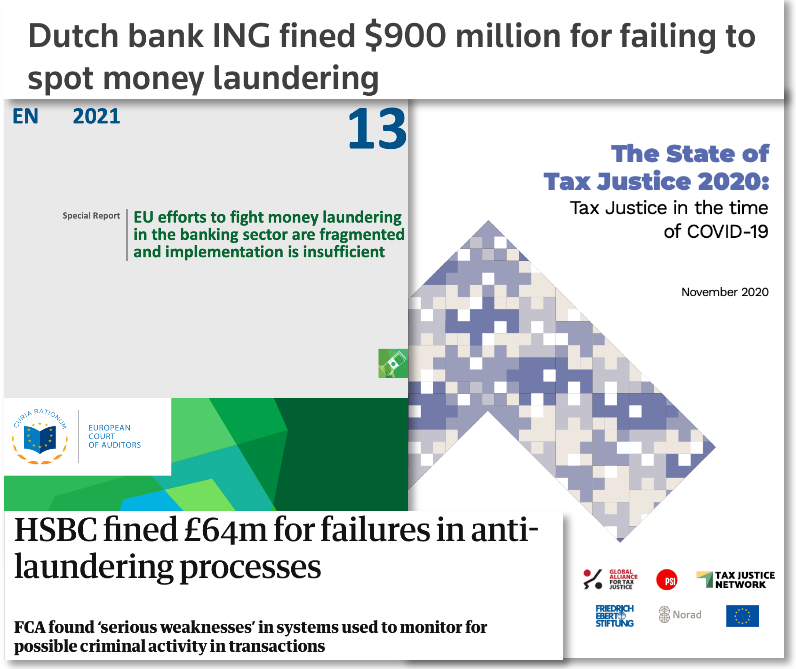
Source: Data Sharing Coalition analysis

Table of Contents

1. **Introduction**
2. Results
3. Potential for other use cases
4. Appendix

Banks can improve detection of financial crime by collaborating on data to create a more complete view of their clients

Banks struggle with financial crime



Dutch bank ING fined $900 million for failing to spot money laundering

EN 2021 13

Special Report

EU efforts to fight money laundering in the banking sector are fragmented and implementation is insufficient

CURIA RATIONUM

EUROPEAN COURT OF AUDITORS

The State of Tax Justice 2020:

Tax Justice in the time of COVID-19

November 2020

GLOBAL ALLIANCE FOR TAX JUSTICE

PSI

TAX JUSTICE NETWORK

FRIEDRICH EBERT STIFTUNG

Norad

HSBC fined £64m for failures in anti-laundering processes

FCA found 'serious weaknesses' in systems used to monitor for possible criminal activity in transactions

Sources: The Guardian, Reuters, European Court of Auditors, World Economic Forum, Tax Justice Network

Data collaboration can improve detection of financial crime

The combat against financial crime is essential to protect society against its harmful effects

- An estimated €2.2 trillion in proceeds from activities such as forced prostitution, terrorism, and drug trafficking are laundered yearly
- The yearly tax losses due to tax evasion are estimated at €387 billion worldwide

Banks struggle with combatting financial crime

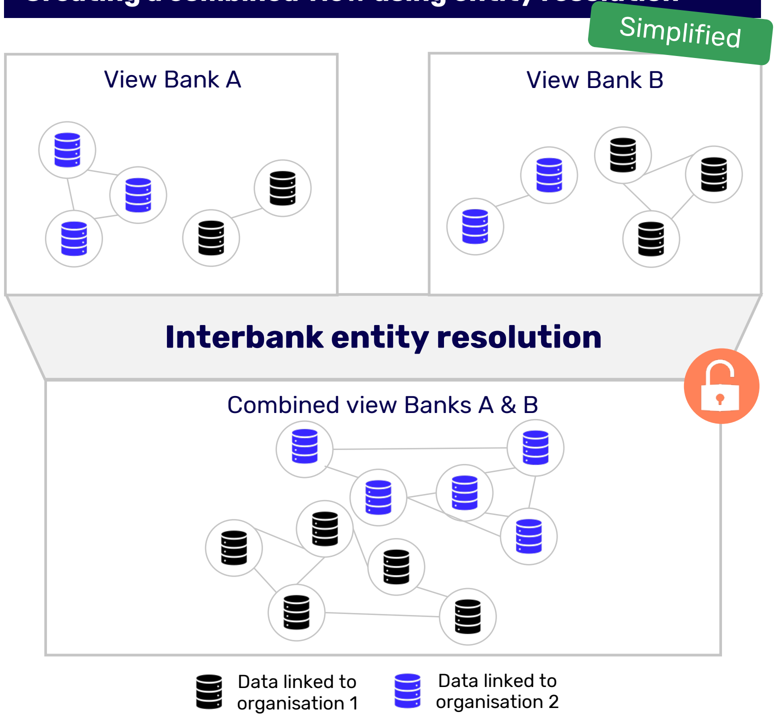
- In order to combat financial crime and comply with AML/CFT regulations, banks need to detect suspicious transactions
- Reportedly, banks have had to scale the size of their compliance departments, significantly increasing operational costs
- However, banks still struggle to detect financial crime. Partly because information is fragmented

A combined view on criminal entities helps banks to improve detection of fraudulent transactions because it reveals complex money flows

- Having insight in the full transaction pattern helps in detecting financial crime
- Criminal entities often operate across multiple banks, setting up complex structures to cover up their activities and resulting in data being fragmented across these banks, making it hard to have a holistic view on transaction patterns
- To detect suspicious transactions, collaborating to combine data across banks would therefore help because it improves insights in the transaction patterns based on which transactions are flagged suspicious

In this use case we explore interbank entity resolution for creating a combined view to improve detection of financial crime

Creating a combined view using entity resolution


Simplified
View Bank A
View Bank B
Interbank entity resolution
Combined view Banks A & B
Data linked to organisation 1
Data linked to organisation 2

Detecting suspicious transactions by creating a combined view

To create a combined view, data from multiple banks needs to be linked to real-world entities (= interbank entity resolution)

- Banks have access only to their own data, this needs to be linked because it often comes from multiple sources (e.g. transaction data, KYC data etc.)
- To create a combined view, available data across different banks needs to be linked to the same real-world organisation
- This process of linking data from different banks is called interbank entity resolution

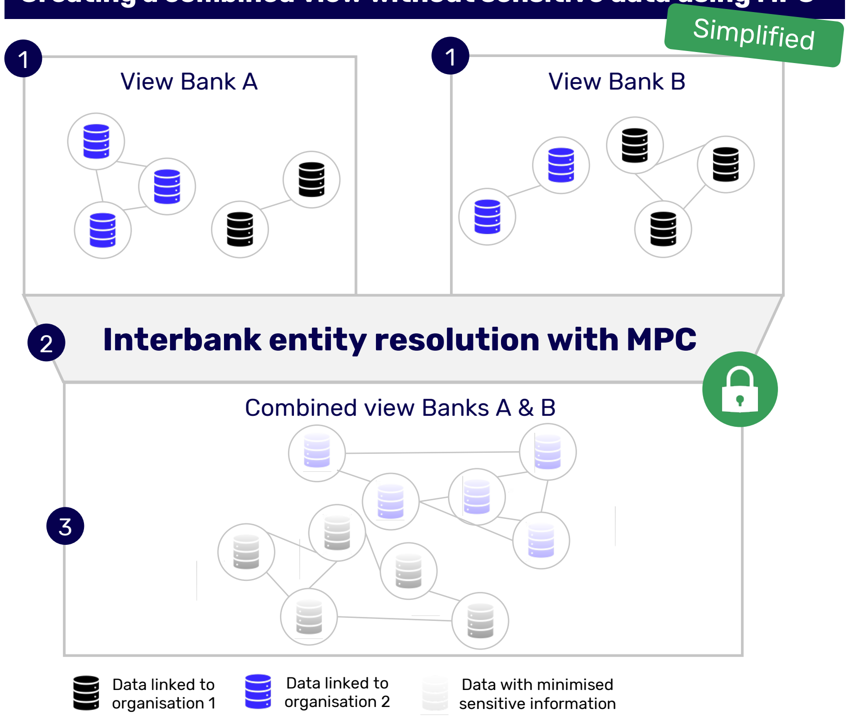
Conducting entity resolution at an interbank level is challenging due to requirements imposed by data sensitivity and confidentiality

- Records from different data sources often lack a common unique identifier, making entity resolution complex
- Entity resolution results can be improved by including sensitive and confidential data in the analysis such as names of the organisation's employees or specific transaction details
- Sensitive and confidential data cannot be easily shared outside the boundaries of the own institution due to restrictions protecting organisations' commercial interests and organisations' and people's privacy
- This creates a challenge for banks that want to collaborate to create a combined view because they cannot use all their data for conducting interbank entity resolution

An illustrative example of how entity resolution can help is included in the appendix p. 18, a detailed process of entity resolution is included on p. 19.

Conducting interbank entity resolution using MPC creates a combined view that contains minimal sensitive data

Creating a combined view without sensitive data using MPC


Simplified
1
View Bank A
1
View Bank B
2
Interbank entity resolution with MPC
3
Combined view Banks A & B
Data linked to organisation 1
Data linked to organisation 2
Data with minimised sensitive information

Using MPC enables including sensitive data without exposing it

- To achieve the best entity resolution results sensitive and confidential data needs to be included in the analysis
- When conducting entity resolution without MPC this means that data needs to be pooled at a central location, exposing it to at least one other party
- Using MPC for entity resolution enables decentralised analysis, meaning that data isn't exposed to parties outside the boundaries of their own organisation

Explanation of entity resolution with MPC

1 The different records (e.g. transaction data, KYC data etc.) reside within each bank. At this step, two options are considered:

- **Pre-resolution**: records are linked to the same real-world entities within each bank (intrabank entity resolution) before serving as an input for interbank entity resolution with MPC
- **No pre-resolution**: records aren't linked within each bank before serving as an input for interbank entity resolution with MPC

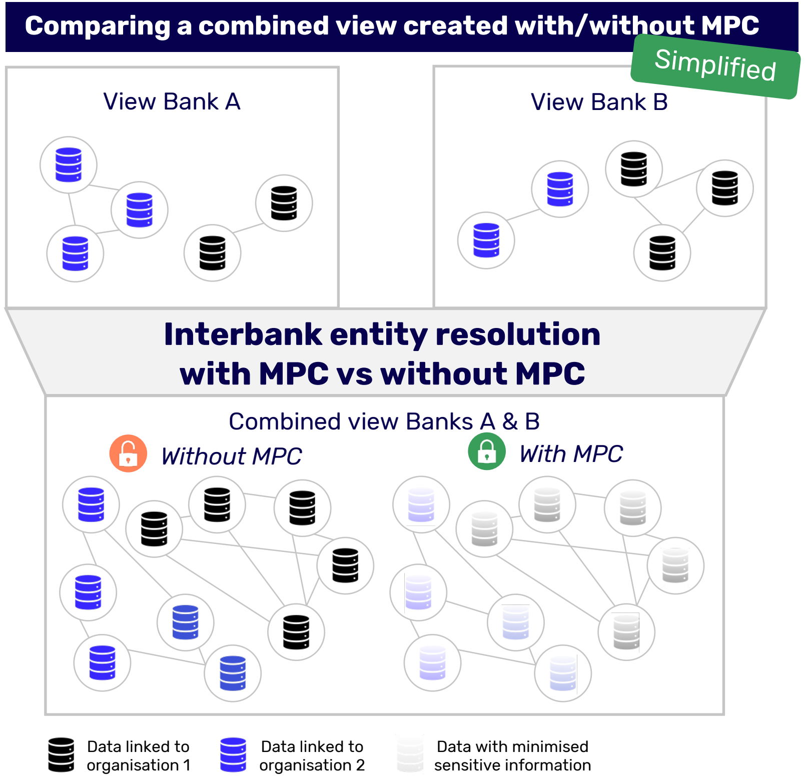
2 Entity resolution with MPC is used to create a view at an interbank level that contains minimal sensitive and confidential information

3 This combined view can be used to detect suspicious transactions at an interbank level without disclosure of sensitive and confidential data to other parties than the one that the data controls

A detailed process of entity resolution with MPC is covered in the appendix (p.20 shows the process with no pre-resolution, and p.21 shows the process with pre-resolution).

In this case study, entity resolution with MPC was compared to entity resolution without MPC to assess its performance

Comparing a combined view created with/without MPC


Simplified
View Bank A
View Bank B
Interbank entity resolution with MPC vs without MPC
Combined view Banks A & B
Without MPC
With MPC
Data linked to organisation 1
Data linked to organisation 2
Data with minimised sensitive information

Assessing potential of entity resolution with MPC

To assess the potential of entity resolution with MPC to create a combined view, the solutions were tested in a business context. The solutions were compared on the following dimensions:

- **Quality**: The difference in quality of the combined view with and without MPC. This relates to the degree to which the solutions are capable of resolving the entities by linking records
- **Scalability**: The degree to which the solutions are able to handle large data volumes. This relates to the computation time that is needed to conduct the analysis and how this relates to the number of records in the dataset
- **Data exposure**: The extent to which sensitive data is shared and banks are in control over their data
- **Allowed processing**: Whether the solution enables to process data as a stream rather than in batches

To test the potential of the solutions with and without MPC samples from the OpenCorporates dataset were used

Background on testing procedures

Dataset used

- To assess entity resolution with and without MPC in the context of detecting financial crime we tested both solutions on a real-word dataset
- A sample from a dataset provided by OpenCorporates[1] was used in this assessment. This data is similar to the data that banks would use to conduct entity resolution in an AML/CFT monitoring context
- The dataset consists of data on both organisations and company directors (officers). Records included in the analysis were:
  - Company names
  - Addresses
  - Postal codes
  - Company officers
- To make a fair comparison, the same data was used to test the solutions

opencorporates

Measurement procedures

Quality: precision and recall

- A sample of ~31K entities was used to test quality of the solutions, this sample consisted of:
  - 11,104 individual entities (i.e. officers)
  - 19,997 organisation entities (i.e. companies)
- Quality tests were performed by running entity resolution without MPC and with MPC (pre-resolution & no pre-resolution), and then comparing the entities left after the analysis to assess precision and recall

Scalability: processing time

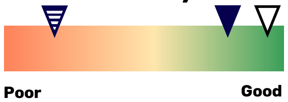
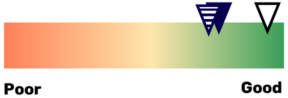
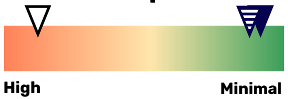
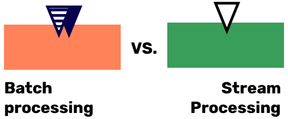
- Two sample sizes were compared to test scalability of the solutions and results were extrapolated to estimate the time to run a larger 10m build
- Scalability tests were performed through increasing the number of records and measuring the processing time required to complete the analysis

Source: 1. OpenCorporates

Table of Contents

1. Introduction
2. **Results**
3. Potential for other use cases
4. Appendix

Entity resolution with MPC performs similarly on quality and scalability and vastly reduces data exposure

Dimensions	Entity resolution without MPC	Entity resolution with MPC	
		No pre-resolution	Pre-resolution
Quality 	**Highest quality** matching as this solution leverages entity resolution in the clear. Therefore, it is used as a baseline for comparison.	**Poor quality** of matching records, as we observed considerably lower recall compared to the baseline.	**Good quality** of matching records, as we observed similar recall and only slightly lower precision (mostly attributable to some over-resolution).
Scalability 	**Best scalability** as computation time scales linearly when more records are added.	**Good scalability** as computation time scales linearly and computations can be run parallel.	**Better scalability** as the solution scales similarly to the no pre-resolution scenario but with fewer comparisons to be made overall runtime is lower.
Data exposure 	**Data exposure is high** because sensitive data has to be pooled in its raw form outside the banks' privacy boundaries to conduct entity resolution.	**Data exposure is minimal** because sensitive data isn't revealed outside the boundaries of the organisation.	**Data exposure is minimal** because sensitive data isn't revealed outside the boundaries of the organisation.
Allowed processing 	**Stream processing is possible** when using entity resolution without MPC, since new increments of data can be added.	**Stream processing is not possible** at this point in time because the current MPC implementation makes use of batch processing.	**Stream processing is not possible** at this point in time because the current MPC implementation makes use of batch processing.

▽ Entity resolution without MPC ▽ Entity resolution with MPC (no pre-resolution) ▼ Entity resolution with MPC (pre-resolution)

MPC hardly compromises the quality of the solution, given that pre-resolution is conducted

Results (N=19,997 Organisations)

Scenario	Pre-resolution (Yes/No)	Entities left after resolution	Resolution compared to baseline	Recall & precision
Entity resolution without MPC	No	2701	-	- -
Entity resolution with MPC	No	17,313	-14,612	0.49 (recall) 0.97 (precision)
Entity resolution with MPC	Yes	2234	+467	0.99 (recall) 0.81 (precision)

How to read these results:
To understand the quality of the entity resolution in the two MPC-scenarios, one needs to differentiate between under- and over-resolution and between recall and precision. What is meant by under-resolution (or low recall) is that fewer entities are resolved than in the dataset which is considered the truth (in this case the baseline), meaning there are many false negatives. Over-resolution (or low-precision) occurs when more entities are resolved than there should have been, meaning there are false positives.
Understanding this detail is important because the difference in number of resolved entities (4th column) does not necessarily tell us how good the resolution is.-Note that the amount of under- and over-resolution can be improved by tuning the entity resolution.

Quality: number of entities resolved is similar

Entity resolution can be done using MPC almost without compromising the quality of the combined view

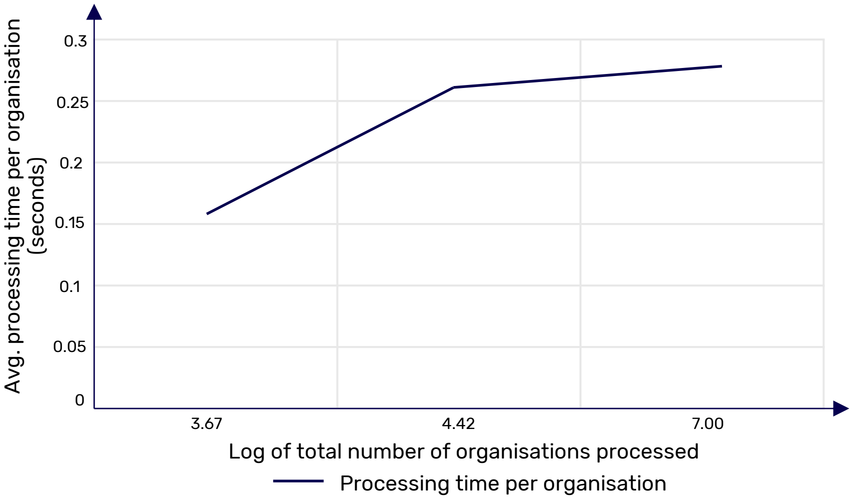
- All scenarios show to reduce the number of entities compared to the original dataset
- Entity resolution with MPC and pre-resolution achieves similar results to entity resolution in the clear when it comes to resolving the number of entities in the dataset
- This implies that the quality of the combined view is compromised only to a limited extent by combining entity resolution with MPC when using pre-resolution

Combining MPC with pre-resolution benefits the quality of the combined view in this use case

- Interbank entity resolution with MPC was conducted both with and without pre-resolving entities within each bank first
- Using MPC without pre-resolution considerably compromises the quality of the combined view compared to the solution with MPC
- Pre-resolution can help overcome the decline in quality when using entity resolution with MPC, although there is some over-resolution happening

Entity resolution with MPC can handle large data volumes due to its linear scaling of computation power required

MPC computation time in relation to number of records



Number of Organisations	Log Number of Organisations
4724	3.67
26,565	4.42
10,000,000	7.00

Scalability: computation time needed

Entity resolution with MPC is a feasible solution to apply within the industry as the computation power that is required to run the solution is manageable

- The required processing time that is needed for the solution with MPC is higher than for the solution without MPC. This is expected because of the extra complexity that comes with the cryptography involved in MPC
- The processing time scales linearly using Roseman Lab's technique for fuzzy matching in MPC
- Adding more MPC servers reduces computation time equally because computations can run in parallel
- The combination of linear scaling and the possibility to run computations in parallel makes the solution feasible within the industry because it is capable of handling large data volumes

Table of Contents

1. Introduction
2. Results
3. **Potential for other use cases**
4. Appendix

Framework to assess whether entity resolution with MPC can benefit a data sharing use case contains 8 factors

Assessment factors for entity resolution with MPC cases			**If factors apply**
1	**Obscurity of records**	Lack of unique identifiers, differences in naming conventions and attributes make it difficult to consolidate records to a real world entity across datasets.	✓ **Suitable use case for Entity Resolution w/ MPC:** There is a strong indication that entity resolution with MPC is specifically beneficial for a number of use cases.
2	**High data fragmentation**	Necessary data is scattered across parties (i.e. multiple organisations are involved).	
3	**High data sensitivity**	Data contains PII and/or is confidential (e.g. because of commercial sensitivity) and thus there are barriers for sharing it for the analysis.	
4	**High complexity of the data**	Data sources that are needed for matching consist of both structured and unstructured data, making the analysis hard.	
5	**High volume of the data**	Data sets that need to be analysed include millions of records.	
6	**High Sector readiness**	Sector participants have the capabilities to implement the solution on a technical, organisational, and governance level.	✓ **Higher chance of long term adoption** If these factors are recognised in a use case, there is an indication that entity resolution with MPC can ensure sustainable value generation for use case participants in a bigger ecosystem.
7	**Social/economic aims**	There is an aim to elevate economic/social challenges (e.g. significantly improve processes, save resources).	
8	**Long term future outlook**	Long term potential (i.e. not a one-time implementation, but is required on a prolonged basis).	

The combination of entity resolution with MPC could benefit a broad scope of use cases beyond combatting financial crime

Example use cases that can benefit from entity resolution with MPC

Indicative

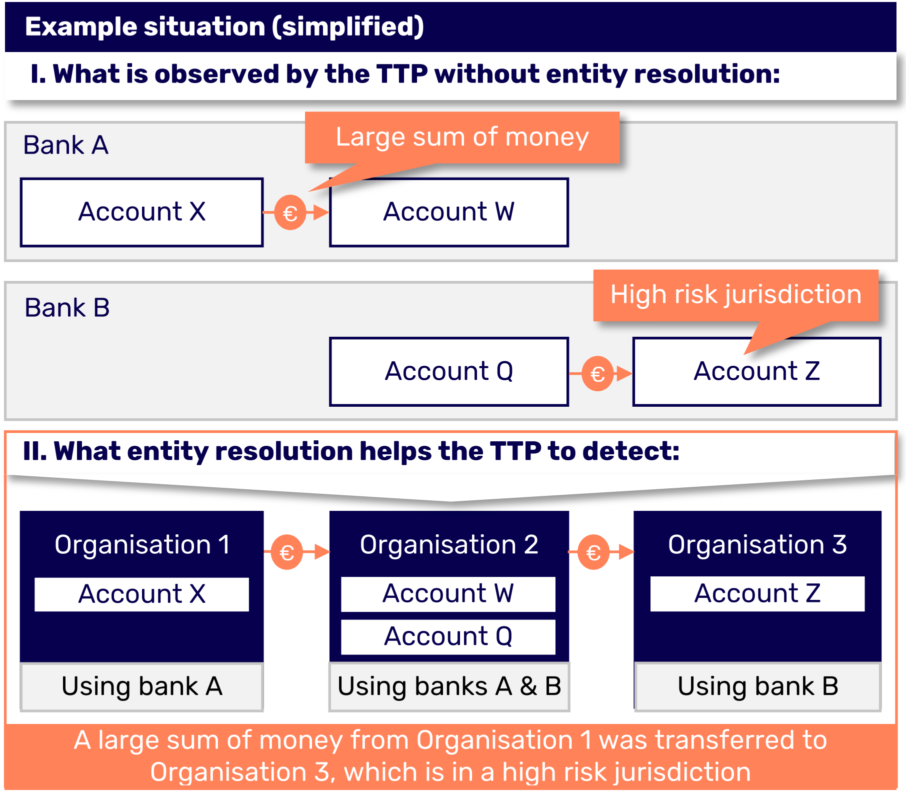
	Monitoring for child protection	Preventing tax evasion	Insurance fraud detection
Context description	Child protection services often miss early stage signals of child abuse due to lack of access to relevant data such as medical reports and notes from teachers[1].	Tax authorities incur losses due to hidden assets and wealth shifting by individuals across borders to lower-tax jurisdictions[2]. Combating tax evasion can be improved through monitoring data sources from various financial institutions (FIs)[3].	In order to detect and prevent insurance fraud, Dutch insurers currently contribute data to a centralised registry[4]. However, the scope of data that is shared this way is limited[5].
Factors observed in the use case*	• **Data obscurity, fragmentation, complexity and volume**: Difficult to link data across institutions involved, using various formats to make records • **Data sensitivity**: Sharing PII data of children is allowed only in special circumstances under GDPR (e.g. serious red flags have been raised) • **Long-term outlook & the aim to elevate social challenges**	• **Data obscurity, fragmentation, complexity and volume**: Data on assets and wealth is fragmented over multiple institutions across borders using different identifiers • **Data sensitivity**: Not all data can be shared directly across borders as legal basis under GDPR applies despite CRS regulation demanding FIs to report to tax authorities • **Long-term outlook & the aim to elevate economic challenges**	• **Data obscurity, fragmentation, and volume**: No unique identifiers (e.g. BSN) are available to match non-health insurance claims to the individual across insurers • **Data sensitivity**: Limited data can be directly shared among insurers due to commercial and privacy reasons • **Long-term outlook & the aim to elevate social challenges**
Value of entity resolution with MPC	Entity resolution with MPC helps to detect abuse in an early stage by generating a combined view, identifying children at risk without revealing underlying sensitive data.	Entity resolution with MPC helps to detect hidden assets and prevent tax evasion by generating a combined view on offshore wealth of an individual across jurisdictions, whilst ensuring data privacy.	Entity resolution with MPC can aid insurers to efficiently generate a combined view on fraudulent insurance claims of an insured individual while preserving data privacy.

Sources: 1. BAE Systems (2019); 2. EconPol (2020); 3. EUTAX Observatory (2021); 4. CIS registry; 5. CIS referencing

Table of Contents

1. Introduction
2. Results
3. Potential for other use cases
4. **Appendix**
    - Detailed example of entity resolution for detection of financial crime
    - Process steps of entity resolution without MPC
    - Process steps of entity resolution with MPC

Introducing a more detailed example of interbank entity resolution helping in the detection of financial crime



Example situation (simplified)

I. What is observed by the TTP without entity resolution:

II. What entity resolution helps the TTP to detect:

A large sum of money from Organisation 1 was transferred to Organisation 3, which is in a high risk jurisdiction

Value of entity resolution for detecting financial crime

- Let's assume three criminal organisations (1, 2 and 3) have set up a chain to cover up transactions. They are using Financial Institution A and Financial Institution B to transact via accounts X, W, Q, and Z:
  - Account X belongs to Organisation 1
  - Accounts W and Q both belong to Organisation 2
  - Account Z belongs to Organisation 3
- A Trusted Third Party (TTP) conducts AML and CFT analyses. For that the TTP applies business rules to detect suspicious financial activity in the transaction patterns
- For example, a transaction is marked suspicious if:
  (a) A large sum of money is transferred; AND
  (b) This money flows into a high risk jurisdiction

I. **The transaction isn't flagged as suspicious in the scenario where interbank entity resolution is not used**:
  - The TTP observes that a high amount of cash has been transferred from account X to account W
  - The TTP knows that account Z is located in the high risk jurisdiction
  - Transactions from X to W and from Q to Z, appear to be unrelated so no flags are raised

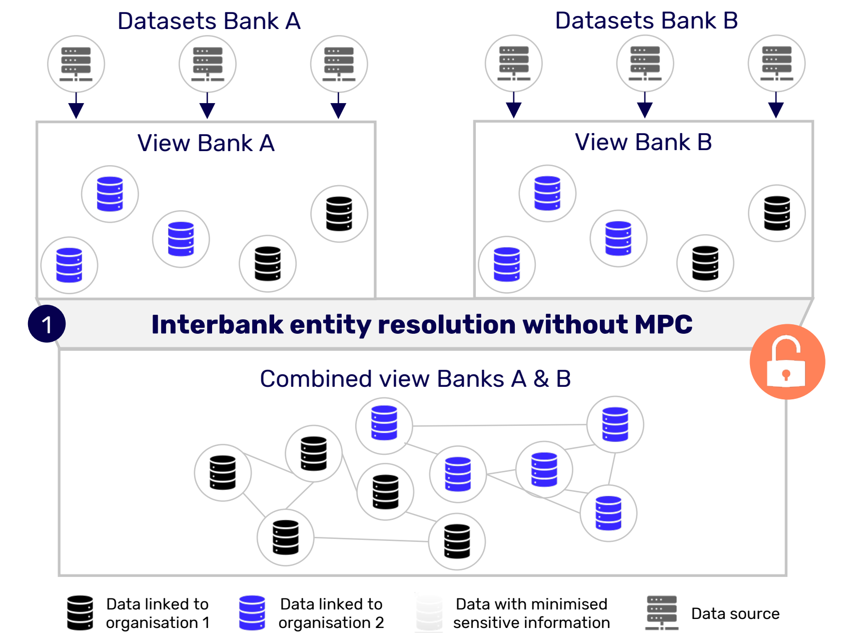
II. **The transaction is flagged when entity resolution is used**:
  - The TTP can see that accounts W and Z belong to the same Organisation 2. And thus, the TTP can track that Organisation 1 has been transferring a large sum of money to an organisation located in a high risk jurisdiction (Organisation 3)

Legend: Bank account | Criminal Entity | Bank | € → Money flow

CoE DSC

Entity resolution without MPC is conducted by pooling all data together and then run the entity resolution analysis

Entity resolution process without MPC


Datasets Bank A
Datasets Bank B
View Bank A
View Bank B
1
Interbank entity resolution without MPC
Combined view Banks A & B
Data linked to organisation 1
Data linked to organisation 2
Data with minimised sensitive information
Data source

Description

A one-step approach is used to arrive at a combined view:

1 Interbank entity resolution without MPC

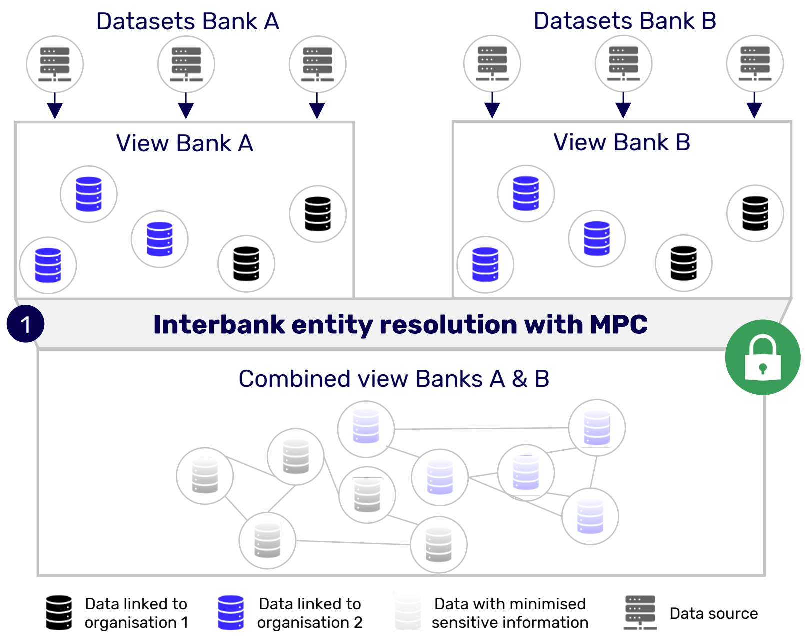
- Data resides at different sources within each bank
- The data is pooled together for the analysis directly for the interbank entity resolution
- This results in the combined view that shows links between records per each resolved entity

Disclaimers:

- Since data needs to be pooled together, it does not remain private

Entity resolution with MPC using one step, where interbank resolution is conducted without pre-resolving the entities

Entity resolution process without MPC


Datasets Bank A
Datasets Bank B
View Bank A
View Bank B
1
Interbank entity resolution with MPC
Combined view Banks A & B
Data linked to organisation 1
Data linked to organisation 2
Data with minimised sensitive information
Data source

Description

A one-step approach is used to arrive at a combined view containing minimised sensitive information.

1 Interbank entity resolution with MPC

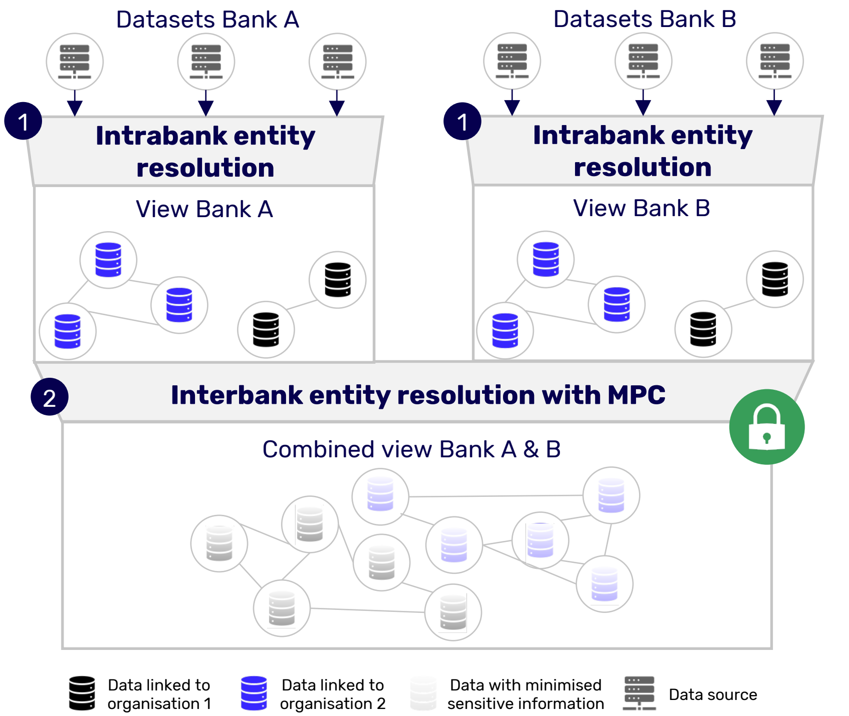
- For running entity resolution with MPC the data is split into buckets
- The buckets are processed using MPC
- As a result, the combined view of resolved entities is generated while underlying data remains private

Disclaimers:

- In this case there is no pre-resolution conducted at each individual bank, entities are directly resolved on the interbank level using MPC

Entity resolution with MPC consisting of two steps where entities are first resolved within each bank before interbank resolution

Two-step entity resolution process with MPC


Datasets Bank A
Datasets Bank B
1
Intrabank entity resolution
1
Intrabank entity resolution
View Bank A
View Bank B
2
Interbank entity resolution with MPC
Combined view Bank A & B
Data linked to organisation 1
Data linked to organisation 2
Data with minimised sensitive information
Data source

Description

A two-step approach is used to arrive at a combined view containing minimised sensitive information:

1 Intrabank entity resolution (no MPC involved)

- Data resides at different sources within each bank
- Intrabank entity resolution is conducted by structuring and analysing the data within each bank
- This process results in a view on all real-world entities within a single bank bank, this first step has 2 benefits:
  1. It improves the quality of the combined view because more information can be used within the boundaries of the bank
  2. It reduces computation time because of reduced complexity (i.e. there are less separate entities after resolution)
- The output of the intrabank entity resolution analysis serves as input for the next step: interbank entity resolution

2 Interbank entity resolution with MPC

- For running the interbank entity resolution with MPC, the data is split into buckets
- MPC processes the buckets in a privacy preserving way
- As a result, the combined view of resolved entities is generated while underlying data remains private