Quick-scan report on the use of

Multiparty Computation

for Mobility Services in the City of Amsterdam

Z. Erkin, T. Turel, F. Geiser

January 2023

Introduction

The Responsible Sensing Lab (RSL) is a collaboration between the AMS-institute and the municipality of Amsterdam that explores how to integrate social values in the design of sensing systems in public space. A part of this exploration consists of researching how to prevent misuse of smart city systems on the level of hardware, software and data governance. Among other approaches, secure multi-party computation (MPC) is a relevant technique in this regard, as it allows the analysis of data of different parties without the need to share the data (Zhao et al., 2019), which leads to an elimination of risks. Therefore, the RSL aims to start a pilot to test MPC in a use case in Amsterdam. In our first exploration of use cases, we focused on mobility, as traffic makes up a large part of the smart city system and it is a domain in which a variety of public and private actors are active, leading to potential situations in which data sharing can be of value.

The goal of the pilot is to investigate the idea of using MPC in a use case within the municipality in the city of Amsterdam. As explained later in the document, MPC is a cryptographic approach that enables computations for a certain function, such as averages, means, with multiple inputs each of which is coming from different parties. These inputs are mostly privacy-sensitive thus, they should be kept secret from the party which performs the computations. Considering that in the field of mobility, the municipality of Amsterdam and all related parties need several types of data from different resources with privacy considerations, MPC can indeed provide a viable solution.

In this report, we first introduce the concept of MPC. We try to address the questions like how MPC works, what are the limitations of MPC, which use cases can be solved using MPC, and what are the differences between MPC and other relevant solutions. Second, we present privacy considerations in the use cases we investigated. These considerations are fundamental as they are defined either by legally or ethically. Third, we provide a list of use cases that we identified based on our interviews with the experts in the field. Fourth, we analyse these use cases and cluster them in one of the groups: a) short-term, b) mid-term , c) long-term project based on criteria such as the number of parties involved, and the type of data resources needed. Finally, we provide advice on how to proceed in two dimensions, namely use cases suitable for development and deployment, and use cases that involve scientific challenges and thus require further research.

Methodology

In the creation of this document, we interviewed several mobility experts and a privacy officer from the municipality of Amsterdam as well as a civil servant from the city of Utrecht. During the interviews, the following points were discussed:

- What are the daily operations that rely on privacy sensitive data?
- What kind of data is required for the daily operations from the municipality and partners?
- What are the privacy requirements legally and ethically?
- What are the desired functionalities of daily and future operations given that the required data is available and there is no privacy risk?

Based on the answers, we identified a number of use cases which we investigate in this report.

Privacy Considerations

In this report, we are focussing on the use cases around mobility where the data points involve the followings:

- Identity of the citizen (e.g. name, BSN, customer ID, etc)
- Age of the citizen
- Identity of the vehicle (e.g. plate number)
- Real time location data (e.g. GPS)
- Speed
- Direction
- Starting point
- Destination point
- Timestamp

In the use cases considered in this report, the data needed for mobility operations should enable better, more efficient and new services to the citizens, while reducing costs and supporting sustainable and environmentally friendly solutions. However, it is legally and ethically vital that these use cases do not create privacy threats for citizens: Identification, tracking and tracing of individuals should be not possible. Furthermore, the use cases should not create business risks for the data providers such as reputation loss or loss of commercially valuable data. In the context of anonymity, the purpose of anonymization is to break the link between the data and its owner. To achieve this goal, there are techniques deployed in practice known as anonymization techniques:

- Suppression of identifiers such as passport number, BSN, names, credit cards, etc,
- Generalisation of data,
- Deploying techniques such as k-anonymization on quasi-identifiers: a combination of attributes in the data set that can uniquely identify the owner. An example of quasi-identifiers is the combination of zip code, gender and age; this combination can identify many individuals.
- Differential privacy for adding noise to the queries such that the existence of a certain individual in a dataset is hidden.

Thus, we assume that the necessary anonymization steps have been already taken, if needed. Our goal is to protect the privacy of the citizens, meaning that no single individual should be identified based on the data processed in the system.

Use Cases

The following use cases were identified:

P1: Targeting a specific audience for crowd management

The city of Amsterdam has a crowd monitoring system with the aim to distribute crowds throughout the city. Tourists, for example, buy Citycards when they are in Amsterdam. The Citycard company thus has data about the tourists that are in Amsterdam. With this data it is possible to target tourists in the city and encourage them to visit certain places and discourage them from visiting others.

P2: Train passenger data from NS

The City of Amsterdam wants to know when it will be busy on the train stations throughout the city. For example, when there is a soccer match or another large event in the city it's helpful to know how many people are coming by train. NS crowd-in-trains data or check-out data of OV-cards can provide such insights. This and other passenger data is not shared. Furthermore, data from the city could be useful for the NS to estimate crowds in the train.

P3: Tram passenger data from GVB

Similarly, the GVB (Bus and Tram) provides hourly data on OV-card check-outs. Amsterdam is interested in more precise data to know more about crowds in the city. Therefore, more refined data on check-outs is more useful. When there are little check-outs however, this data gets more privacy sensitive.

P4: Smart City sensors

The coalition agreement of the City of Amsterdam states that it should be possible to move through the city without being spied upon. However, with the growing amount of cameras and sensors in the city undermine this goal. The mobility department has the largest number of sensors in the city. However, it is not possible to use the sensory data to the fullest extent due to privacy considerations. Under privacy guarantees, the sensor data can be utilised for use cases in the city without having access to the raw data, decreasing the amount of data that the city gathers about citizens in public space.

P5: Combining different data sources to determine origin destination relations

The municipality is interested in improving traffic management. Knowing more about the origins and destinations of drivers in the city is key to refine road level interventions. One way to gather this data so is to scan the licence plates of cars with the different licence plate camera systems in the city; Intelligent Access, scan cars of parking control, parking garages etc. With privacy guarantees, these relations can be analysed without sharing licence plate information and accessing the personal data of drivers.

P6: Sharing mobility data between government agencies

Other government agencies such as the Provincie Noord-Holland or Rijkswaterstaat also have data about traffic from their cameras. It is useful for the city of Amsterdam to have insights from this data. Combining these different sources can enable better and new services.

P7: Sharing GPS data

For some processes in the city, it is necessary to know the GPS location of certain parties in the city. For example, to give police cars priority at intelligent traffic lights based on GPS location. However, GPS-locations of such vehicles are sensitive data.

P8: Getting information from companies

Many companies in the city have mobility data. New flash delivery services cause additional bike traffic. Knowing where their dark stores are and their biking routes give extra insight on their mobility patterns, which is interesting for the city. However, this is sensitive data that companies do not want to share, let alone in this competitive business. With privacy protection, the data of these companies can be analysed to gain insights without the city or competitors having access to it. Examples of data are the amount of rides, the reason for the rides and which things are transported. It is also an option to only give permits to these companies if they provide their data under privacy guarantees.

P9: Optimising the empty running of delivery services

Another use case concerns parcel delivery. Now the different parcel delivery companies all have their own vans, which is inefficient. Often vans of different companies can be found in the same streets, causing a hindrance of traffic. Collaboration and load sharing based on data about the loads, amount of vans and routes they drive shared between these companies makes it possible to maximise load capacity and use these vans more efficiently. This can lessen the nuisance caused by these vans in the city. However, this is sensitive data that companies are not willing to share. With privacy guarantees, insights can be gained to optimise loading without giving up any data.

P10: Enabling specific options within navigation systems (exemptions, permits, limited traffic zones, vehicle properties etc)

Whilst driving it is sometimes convenient to have certain information at hand such as exemptions to certain areas and permits of the driver. One way to offer this information is through navigation systems such as Google Maps or TomTom. To do so, however, would require to offer quite some personal information to these companies that are connected to the exemptions, permits, etc..

P11: Optimising the network together with mobility providers in the city

There are many businesses that offer mobility services in the city. More businesses means more vehicles that take up space in the city. When these mobility providers share data about the use of their services, the offer of vehicles can be made more efficient and the amount of vehicles in the city reduced. For example, data about where sharing scooters or bikes are located, how often they are used etc. However, this is sensitive data which is currently not shared with the city. With MPC, this data can be processed for further services, e.g. optimisations and analysis, without the city or competitors having access to them.

P12: Pay according to use

The national government announced plans to introduce road pricing per 2030. The city of Amsterdam may decide to start working towards such a system in the near future. It is interesting for a city to differentiate between different drivers. For example, between residents and visitors or other user types. This requires all kinds of data on driving which contain personal information.

P13: Creating an even playing field by providing open data

Today, businesses that have the most data can offer the best services. This creates a 'data monopoly', in which it is difficult for new parties to enter the market. If businesses are required to share their data this would create an even playing field. Businesses do not want to share data, however, due to its sensitivity. With privacy guarantees and the right regulations, data from businesses can be open for the market without sensitive information being shared.

P14: Detecting car plates for parking penalties

There are several cameras and sensors installed throughout the streets. The sensory data can be used to detect cars which are parked illegally. However, such a system poses significant privacy risks. Designing such a system with privacy guarantees, however, will provide significant advantages to the city.

P15 De Digitale Gracht

The Digital gracht is a system that is used to enforce local regulations on noise and speed on the waterways of Amsterdam. It consists of various sensor systems (AIS beacons, RFID sensors for vignettes, noise sensors and smart cameras) inform a central dashboard.

Not all data on boats needs to be available to all users at all times. For instance: in the case of a boat sinking, the data about its user should be accessible, but not otherwise.

Privacy by Design

Privacy protection is an important goal today, particularly after GDPR entered into force in 2018. Public awareness is also increasing due to numerous privacy scandals around popular services. What is more, several organisations and companies see the advantage of collaboration around sharing data for more accurate services, better planning and new business opportunities. We see many attempts in this direction including but not limited to supply chain logistics, combat against financial crime and medical data sharing across the EU.
It is though essential to protect the citizen against potential privacy risks. For this purpose, there are a number of guidelines such as Privacy by Design principles. Hoepman (2018) provides 7 steps for privacy protection that are minimise, separate, abstract, hide, inform, control, enforce and demonstrate. In this work, we are focussing on collaboration and protection of sensitive data to disable the identification of a single person based on the processes data and thus, multi-party computation in a decentralised setting as a part of "separation" is our focus.

Multiparty Computation for Privacy

Secure multi-party computation is a cryptography approach presented by Yao (1982). The main idea is that any function with multiple secret inputs from different parties can be computed securely. To illustrate this idea, let us consider a toy example. Imagine that a group of students in a classroom want to calculate the average age without telling their exact age. Here, the function is the average, which requires the division of total age by the number of students. The secret inputs are the ages from each student, which should be kept hidden from all students.

Secure computation of average age

For the example above, consider that we want to use pen and paper. Each student writes down his or her age on a piece of paper and folds it. Here, it is essential that the shape of each paper and the writing of the students are identical and cannot be distinguished. A trusted person, let say the teacher, collects the papers and in front of the students, opens the papers and tallies the ages. Finally, the teacher announces the average age.
The above simple protocol is effective as it does not leak the secret inputs under certain conditions: The papers and the writings are identical, the teacher is trustworthy and the students really report their actual age. However, in real life, we cannot devise protocols based on pen and paper such as this one. Nevertheless, the approaches in cryptography follow the same principles.

Approach 1: The presence of a trusted entity.
In the case of having a trusted third party (TTP), the computations on secret inputs are relatively simple. Imagine that an organisation, e.g. a governmental body or non-profit organisation that can be trusted by many other organisations, can collect the data from several other organisations and perform the required operations on them. During the transmission of the data, proper security measures such as encryption should be deployed.
This approach with a TTP is the ideal solution as it does not leak privacy-sensitive information. However, having a TTP is not easy, and in many cases not possible. Even in the case of its presence, maintaining a TTP is costly.

Approach 2: MPC based on additive secret sharing

In this approach, we rely on cryptographic constructions. An important remark here is the concept of random shares. Imagine that we have a secret value: 42. We want to share this value among three parties: Alice, Bob and Charles. Alice is given her random share 15, Bob is given his random share 18 and Charles is given: 42-15-18=9. Each person sees a random number that does nor provide any insight about the secret value 42. There is only one way to reconstruct the secret: by combining all the secret shares. Imagine that another secret value, 87 is also shared among Alice, Bob and Charles, as follows: Alice has 5, Bob has 27 and Charles has 55. Table 1 shows an overview of the values used in the examples.

Given these secret shares, Alice, Bob and Charles easily compute the desired function of 42 and 87. For example, addition and subtraction of these numbers by using only their shares. In case, they want to multiply 42 and 87 using their shares, it is also possible but requires an interactive step that we skip in this document.

Recall that basic operations such as addition, subtraction and multiplication can be used to construct more advanced operations such as divisions and comparisons. Once we have all these operations, any function or algorithm can be designed based on secret shares.

This approach, as long as the shares are created (pseudo)-randomly, is very secure. In cryptographic terms, it provides perfect security and even quantum computers cannot obtain the secret value since every value has the same probability of being the correct one. MPC based on additive secret sharing is a well-studied topic in academia and there are numerous research articles on how to design a certain algorithm.

The drawbacks of the MPC approach based on additive secret sharing can be summarised as follows:

- A trusted dealer is needed to create the shares initially. This dealer should destroy the secret after the creation and disappear.
- Addition and subtraction are easy computations and can be performed locally with no interaction. Multiplication, however, is an interactive operation.
- Due to the interactive nature of the required operations, the bandwidth requirement is high.
- Even though threshold versions can be designed, the generic MPC requires the parties to be online for the successful completion of the operations.
- It is assumed that parties do not collude to obtain the secrets.

Table 1: Two examples of MPC based on additive secret sharing.

	Secret share of 42	**Secret shares of 87**
Alice	15	5
Bob	18	27
Charles	9	55
Secret is the total:	**42**	**87**

Approach 3: Secure computation based on homomorphic encryption schemes

Certain encryption schemes preserve some structure after encryption such that the plaintext data can be manipulated under encryption. For example, given the encryption of two messages E(a) and E(b), it is possible to obtain the encryption of their sum, product, or both by simply operating on their ciphertext:

$$E(a + b) = E(a) \times E(b)$$
$$E(a \times b) = E(a) \text{ oplus } E(b).$$

Paillier's (1999) system is additive thus, providing only the addition of plaintext values but not their product. However, using custom protocols, it is possible to design protocols for multiplication, division and comparison. These protocols like MPC based on additive shares are interactive, meaning there is a need for communicating with the decryption key holder.

Imagine that Alice, Bob and Charles want to calculate the average age as before. Dave holds the decryption key. Under this assumption, we can create a protocol using an additively homomorphic encryption scheme as follows:

1. Alice, Bob and Charles receive random numbers r1, r2 and r3 such that r1+r2+r3=0.
2. Alice encrypts her age plus r1: E(A+r1) using the encryption key of Dave and sends it to him,
3. Bob computes E(B+r2) and sends it to Dave,
4. Charles does similarly and sends E(C+r3) to Dave.
5. Dave, using the homomorphism property, calculates the encrypted sum E(A+B+C+r1+r2+r3) and decrypts it. The result will be the total age, since random values will cancel out.

For the simple protocol above, Dave cannot see the ages of any other party other than the total age. This is guaranteed by the random values used in the protocol. However, in the case of more sophisticated computations, Alice, Bob and Charles need to work with Dave in an interactive manner, increasing the computational and communication overhead.

Gentry (2009) proposed the first fully homomorphic encryption scheme that is both additive and multiplicative. This breakthrough enabled us to compute the sums and products without interaction. Currently, that encryption scheme, its variants and similar others are still not practical enough to use in practice: a single encryption can take up to seconds to complete and the key lengths for the encryption schemes are too large. However, there is significant progress in the efficiency of these schemes.

Approach 4: Hybrid Approach

Depending on the computations and the number of entities, it is possible to design a system using one of the above approaches. In many cases, the research challenge is to achieve efficiency in terms of run-time speed, bandwidth and storage. Therefore, research articles provide custom designs per challenge. A number of articles also provide hybrid solutions, that is designing a part of the protocol using one approach and the other part with another approach. Which approaches can be used in combination heavily relies on the application setting and privacy requirements.

Malicious Stakeholders

In MPC, the design of the protocol relies on certain assumptions such as the behaviour of involved stakeholders. There are two commonly used models for behaviour: semi-honest and malicious models.

Semi-honest security model assumes that the involved stakeholders are honest and thus, follow the protocol steps properly. However, they are also curious so that they collect publicly available data and try to learn more than the protocol defines. Hence, the model is also known as an honest-but-curious model.

Another model assumes that all stakeholders are malicious: they can provide incorrect inputs, perform incorrect computation or no computation to deviate from the protocol description. Hence, the model is known as a malicious model.

In literature, a significant portion of the proposed solutions are based on a semi-honest model. The switch from this model to the malicious model is by deploying extra steps for checking input and computation correctness, e.g. using commitment schemes and zero-knowledge proofs. It is important to note that these checks introduce more overhead in terms of computation and communication.

Analysis of the Use cases

The use cases investigated in this work and the related data source for each of them are given in Table 2. The table also provides the main objective of each use case.

A common denominator for each use case is the sensory data as seen in Table 3: this could be GPS location, camera input, noise-level detector, speed and direction of a vehicle. The information needed for each use case based on these sensory data can be clustered in two groups:

1. Explicit information: statistical analysis.
2. Implicit information: behavioural analysis.

For the first group, statistical analysis can be achieved by observing the available data. For the second group, a more detailed study needs to be designed that also involves experts from different fields such as psychology. However, for both groups, we see the following common points:

- There are multiple data sources needed to achieve the goal,
- Data owners have legal or business related concerns to collaborate,
- There are strong incentives for collaboration in terms of better planning, optimisation of existing services and new services.

Given the nature of the use cases, using multi-party computation techniques for explicit and implicit information gathering is a solution that can satisfy legal and commercial concerns. On one hand involved stakeholders can protect their commercially valuable data, on the other hand they can benefit from the advantages of collaboration. Therefore, we advise to develop a prototype based on a use case to demonstrate the usability of MPC. However, there are also scientific challenges. as well as governance challenges that will have an impact on the technological solution. In the followings, we highlight a few of these challenges:

- Involved stakeholders and their roles: There are some stakeholders that have data and business interest. The Amsterdam municipality also has a role as the stakeholder that wants to make decisions. Involving, non-profit and non-governmental organisations in the design can significantly increase the trust factor.
- Efficiency: It is essential to identify the efficiency requirements in terms of run-time, bandwidth and storage of the privacy-pre-

serving version of the use case since the proposed solution will introduce overhead. In other words, the design for a real-time monitoring system is different from a statistical analysis system that works in certain time intervals.
- Security challenges: The scientific solution will rely on the security requirements such as access control, auditing, dispute resolving, key management and honest or malicious behaviour of the stakeholders.

The authors of this report advises the following steps:

1. Identification of a use case with explicit statistical data analysis,
2. Establish a small consortium that can provide data and use case details,
3. Design of the privacy-preserving system using MPC,
4. Implementation of the design as a prototype.

We also recommend involving non-profit and non-governmental organisations to investigate the use case for privacy, legal and ethical aspects.

Table 2: Use case classification based on the data they rely on.

Use Case	Data Source	Objective
P1	Citycard	Time-dependent suggestions
P2	NS and Translink	Statistics and suggestions
P3	GVB	Statistics and suggestions
P4	Cameras and sensors	Statistics and suggestions
P5	Car plate and location data from cameras	Low emission check Parking fees Input for policymakers
P6	Traffic data from cameras	Input for policymakers
P7	GPS location data	Intelligent traffic guidance
P8	GPS location data of bikes/scooters	Optimisation of services Input for policymakers
P9	GPS location and transportation capacity	Load balancing for transportation

P10	GPS location and municipal data	Customised navigation
P11	GPS location data of commercial bikes and scooters	Optimisation of services Input for policymakers
P12	GPS location and personal data	Pay per km
P13	Any data	Data Sharing
P14	GPS location, personal data and car plate number	Data Sharing
P15	GPS and other sensory data	Data Sharing

Table 3: Classification of data types, discrete or continuous location data, per use case.

Location Data	**Individuals**	**Vehicles**
Discrete Point (Check in/out)	P1, P2, P3, and P4	P5 and P6
Continuous (GPS)	-	P7, P8, P9, P10, P11, and P12, P13,P14, P15

References

Gentry, C., (2009). Fully homomorphic encryption using ideal lattices. In Proceedings of the forty-first annual ACM symposium on Theory of computing (STOC ‘09). Association for Computing Machinery, New York, NY, USA, 169–178. https://doi.org/10.1145/1536414.1536440

Hoepman, J-H., (2018). Privacy Design Strategies (The Little Blue Book). Creative Commons Attribution - Non Commercial 4.0 International License (CC BY-NC 4.0).

Paillier, P., (1999). Public-key cryptosystems based on composite degree residuosity classes. In Proceedings of the 17th international conference on Theory and application of cryptographic techniques (EUROCRYPT’99). Springer-Verlag, Berlin, Heidelberg, 223–238. https://doi.org/10.1007/3-540-48910-X_16

Yao, A.C., (1982). Protocols for secure computations. 23rd Annual Symposium on Foundations of Computer Science (sfcs 1982), pp. 160-164. doi: 10.1109/SFCS.1982.38.

Zhao, C., Zhao, S., Zhao, M., Chen, Z., Gao, C., Li, H., & Tan, Y. (2019). Secure Multi-Party Computation: Theory, practice and applications. Inf. Sci., 476, 357-372. https://doi.org/10.1016/j.ins.2018.10.024.

Responsible Sensing Lab

Technologies like smart sensors can help solve urban challenges. But when collecting data, what public values are involved? The Responsible Sensing Lab explores how to integrate social values in the design of sensing systems in public space.

responsiblesensinglab.org   /responsiblesensinglab