Multi-Party Computation in the GDPR

Lukas Helminger[1,2] and Christian Rechberger[2]

[1] Know-Center GmbH, Graz, Austria
[2] Graz University of Technology, Graz, Austria
{lukas,christian}@iaik.tugraz.at

Abstract. The EU GDPR has two main goals: Protecting individuals from personal data abuse and simplifying the free movement of personal data. Privacy-enhancing technologies promise to fulfill both goals simultaneously. A particularly effective and versatile technology solution is multi-party computation (MPC). It allows protecting data during a computation involving multiple parties.
This paper aims for a better understanding of the role of MPC in the GDPR. Although MPC is relatively mature, little research was dedicated to its GDPR compliance. First, we try to give an understanding of MPC for legal scholars and policymakers. Then, we examine the GDPR relevant provisions regarding MPC with a technical audience in mind. Finally, we devise a test that can assess the impact of a given MPC solution with regard to the GDPR.
The test consists of several questions, which a controller can answer without the help of a technical or legal expert. Going through the questions will classify the MPC solution as (1) a means of avoiding the GDPR, (2) Data Protection by Design, or (3) having no legal benefits. Two concrete case studies should provide a blueprint on how to apply the test. We hope that this work also contributes to an interdisciplinary discussion of MPC certification and standardization.

Keywords: Multi-Party Computation · GDPR · Compliance · Privacy Enhancing Technologies · Privacy by Design

1 Introduction

The EU General Data Protection Regulation's (GDPR) [19] two primary objectives seem to interfere with each other. On the one side, the GDPR is best known as the world's strictest privacy and security law. So the majority believes it is all about protecting individuals from personal data abuse. On the other not so well-known side, the GDPR aims to simplify the free movement of personal data. In that sense, the regulation is also very business-friendly. Obviously, there are situations where those two aims create tension.

Privacy-enhancing technologies (PETs) promise to manage the balancing act between personal data protection and open data economy, at least to some extent. The idea of using technology for data protection whilst simultaneously not restring data-driven business is rooted in Article 25 GDPR (Data protection

by design and by default). For a data controller, the possible range to choose a suitable PET is extremely broad. It ranges from access controls over VPNs to cryptographic concepts like multi-party computation (MPC) [36], differential privacy [15], and zero-knowledge proofs [22]. Especially technologies that reduce the need to trust data controllers have received a lot of funding lately. As a result, one could see a high level of research activity in PETs in the last couple of years. Fortunately, this effort led to significant performance gains up to a point where the technologies are scalable to enterprise-size data sets. Also, the technical readiness level improved so that more and more PETs can be called state-of-the-art. Consequently, the research results were exploited, and today many companies are offering advanced PETs.

Nevertheless, the adoption of PETs leaves a lot to be desired. In particular, the most effective PETs are being neglected to a large extend. Instead, businesses and public authorities choose organizational measures and weak PETs. A major reason for that is the legal uncertainty involving PETs. Often the impact of a specific PET regarding GDPR compliance is hard to estimate. It is only understandable that businesses want to know the legal consequences of using a PET before deploying it.

The purpose of this paper is to assess the GDPR compliance of multi-party computation. MPC is a highly versatile PET. Its application ranges from secure distributed genome analysis [26] over collaborative fraud detection [31] to privacy-preserving machine learning [27]. The ability to protect data during a joint computation involving several parties makes it an excellent fit for the GDPR. It drastically minimizes the need to trust data controllers or processors. In addition, MPC facilitates privacy-friendly data sharing across different companies. Lastly, as a subfield of cryptography, MPC comes with mathematical guarantees. The security level can be compared to long-established encryption standards.

Paper Organization. In Section 2, we provide an introduction to MPC. Due to the lack of space, we omit all technical details that are irrelevant from a legal perspective. We then focus on the GDPR articles and their interpretations that are of particular concern to MPC. More concretely, we try to understand the definition of personal data (Section 3) and the meaning of data protection by design (Section 4). In Section 5, we present a test for assessing the legal implications of the use of MPC. In addition, we apply our test to concrete use cases in Section 6.

2 Multi-Party Computation

This section aims to provide an overview of MPC. We focus on the aspects relevant to the GDPR and try to refrain from being unnecessarily technical. For a more technical rigorous treatment of MPC, the reader is referred to these excellent textbooks [20,32,10]. Before starting with MPC, we want to recall what generally is understood by the term computation (alternatively also algorithm

or analysis). A computation is a procedure transferring some given input data to an output.

2.1 Introduction

MPC is a subfield of cryptography dating back to the 1980s. It allows two or more parties to input data and receive output in a privacy-preserving way. In particular, MPC is able to protect the input of each party during a mutual computation. The data protection guarantees offered by MPC can be best understood by the following thought experiment. Imagine an ideal world where there exists a fully trusted third party recognized by everyone. Then, whenever two or more distrusting parties want to analyze their combined data, they send their data to this trusted third party. It then performs the requested analysis on the pooled data and returns the result (output) to the parties. Since the trusted third party cannot be corrupted, nothing except the output of the computation gets revealed to the parties.

MPC replaces the need for such a hypothetical trusted third party by means of cryptography. In other words, advanced encryption-like techniques provide the same data protection in the real world as the trusted third party in the ideal world. To sum up, MPC protects input data during computation but not the computation's output.

2.2 Private Set Intersection

The introduction of a specific MPC protocol[3] called private set intersection (PSI) [13] will offer an intuitive view of the concept of MPC. PSI allows two parties to compute the intersection of their data sets jointly. Thereby, neither party learns information from the protocol except for the elements in the intersection. Consider two hypothetical companies that would like to find out their shared customers. However, both are reluctant to share their list of customers. PSI enables them to perform a seemingly counterintuitive computation, where both end up knowing the names of their shared customers but nothing more (see Figure 1).

Instead of sending both data sets to a trusted third party, a PSI protocol will encrypt both data sets. Then, the whole computation (calculating the intersection) is performed on encrypted data shares. Therefore, the input from every part is protected and cannot be retrieved by the other party. This line of argumentation applies to every MPC protocol and is essential for the classification of MPC in the GDPR.

2.3 Security

There are three dimensions to MPC security. The academic community considers MPC protocols to be secure if they offer at least 128-bit computational security

[3] If we write MPC protocol, we mean a specific MPC (program would be a synonym for protocol).

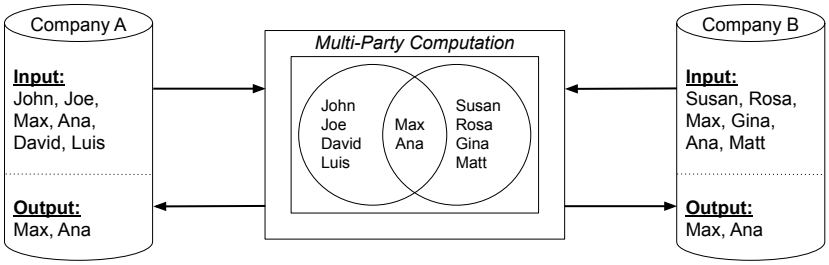

Fig. 1. Private Set Intersection

(guards against brute-force attacks). This is equivalent to AES security [11], the most popular encryption scheme and often used as a benchmark for security level.

The second dimension is the so-called security model of the MPC protocol. There are two major ones the semi-honest and the malicious security model. In the semi-honest security model, it is assumed that no party deviates from the protocol. More concretely, security is guaranteed as long as every party follows the protocol. In contrast, the malicious security model protects the input data even if some party deviates from the protocol.

Orthogonal to the security model is the trust assumption. Any MPC protocol implicitly makes a trust assumption. Protocols differ on how many colluders can be tolerated before security is broken.

3 GDPR: Personal Data

The GDPR only applies if personal data is processed, whereas non-personal data falls outside its scope of application. Thus the definition of personal data is of uttermost practical relevance to anyone processing data. Article 4(1) GDPR defines personal data as follows:

> 'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

We can already see that this definition is very broad. There is one passage in the GDPR's preamble, a so-called recital that aims to clarify the definition of personal data. Although such recitals are not legally binding, they provide information about how the articles should be read and are often considered in

court. Recital 26 GDPR tries to offer a legal test to differentiate between personal and non-personal data.

> To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.

There are two widely recognized interpretations of Article 4(1) and Recital 26 GDPR. Because their differences impact how MPC is seen in the GDPR, we summarize the main arguments.

3.1 Absolute Approach

The absolute approach argues that as long as a data object is personal data to someone, it is personal data to everyone. In other words, personal data is independent of the perspective. This even holds if the data necessary to identify a data subject is spread between different parties.

Note that we are only talking about the absolute approach in terms of perspective [16]. In the literature, sometimes, the absolute approach is mixed with the questions of whether the GDPR favors a risk-based approach. Although there are links between the discussion of the risk and the perspective, we want to keep them separate for clarity (see Section 3.3).

The main argument in favor of the absolute approach is the wording 'by the controller or any other person' in Recital 26 GDPR. The most authoritative document emphasizing the absolute approach is the Article 29 Working Party's (WP29) opinion [2] (now the European Data Protection Board). It states that rendering personal data anonymous should be as permanent as erasure and thereby achieve irreversible de-identification. To further clarify the implications on practical matters, they provide the following cautionary example. A data controller collects data on individual travel movements. Then the data controller removes direct identifiers and subsequently offers this data set to third parties. WP29 concludes that even if nobody except the data controller could identify a data subject in the dataset, the data would be personal data to everyone.

National authorities' rulings further legitimate the absolute approach. As pointed out by Bergauer et al. [3], the Austrian Data Protection Authority considered that data can still qualify as personal data despite the fact that the data controller itself cannot identify a data subject[4]. Similarly, the French Conseil d'Etat - the highest national administrative court - emphasizes that there is no difference between whether the data subject can be identified by the data controller or a third person[5].

[4] DSB-D122.970/0004-DSB/2019.
[5] ECLI:FR:CECHR:2017:393714.20170208.

3.2 Relative Approach

In contrast, the relative approach argues that personal data is a relative concept. More specifically, it is sufficient to only look at the data controller's perspective to determine whether information constitutes personal data. As a consequence, the same data item can be anonymous for one party while being personal data for another party.

Advocates of the relative approach stress that the emphasis should lie on the term 'means reasonably likely to be used'. According to them, the logic of the case Breyer v Germany[6] favors the relative approach.

The judgment of the European Court of Justice (ECJ) in Breyer v Germany is the leading case on the interpretation of Recital 26 GDPR. Breyer's dynamic IP address was collected by the German government when he visited a public authority website. The government could not identify him without additional information held by an Internet Service Provider (ISP). The ECJ concluded that the IP address qualified as personal data because of the government's power to obtain the additional information from the ISP in the event of a cyberattack. Mourby et al. [28] draw the conclusion that in the absence of this legal channel, the IP address would not have been personal data for the government. Thus confirming that personal data is a relative concept.

3.3 Risk-Based Approach

There is a broad consensus that Recital 26 GDPR formulates a risk-based approach. More concretely, if identification is not reasonably likely, data can be considered non-personal. Recital 26 GDPR states factors which shall be taken into account in such a risk assessment:

> To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.

If the GDPR would not favor a risk-based approach - i.e., a zero-risk threshold - then every data would account for personal data. There is always at least a theoretical possibility that data can be linked to a natural person. For instance, even sensor data could give information about the person that installed the sensor.

Important authoritative documents clearly reject a zero-risk threshold. WP29 points out that 'a mere hypothetical possibility to single out the individual is not enough to consider the person as identifiable' [2]. Moreover, the Irish Data Protection Authority argues that: '[I]f it can be shown that it is unlikely that a data subject will be identified given the circumstances of the individual case and the state of technology, the data can be considered anonymous' [8].

[6] Case C-582/14 Patrick Breyer [2016] EU:C:2016:779.

We stress here that, in our opinion, the risk-based approach is orthogonal to the discussion relative versus absolute approach. The risk-based approach can equally be applied to the absolute and the relative approach. It only differs what perspective the data controller has to take into account in the risk assessment. In contrast, advocates of the relative approach often use the risk-based approach as an argument in their direction [21]. They are reasoning that a data controller cannot sensible calculate the risk of identification from the perspective of every party in the world.

3.4 Conclusion

We want to summarize the implications of the different presented approaches to this work. First, any analysis of PETs in the context of the GDPR makes only sense if there is not a zero-risk threshold. No technology can offer a 100% guarantee. So we have to assume some risk tolerance in our analysis. Because of the uncertainty regarding the absolute versus relative approach, our analysis covers both. The use cases in Section 6 should highlight the differences in the approaches concerning MPC.

4 GDPR: Data Protection by Design

Cavoukian, back then Ontario Privacy Commissioner, coined the term "Privacy by Design" (PbD) in 2009 [5]. PbD aims to integrate privacy objectives into the entire development of personal data processing technology. In contrast to the prevalent design process, where privacy concerns are discussed in a late stage of the development, if at all. In addition, PbD is about organizational procedures that enhance privacy during personal data processing. PbD is an interdisciplinary concept combining law and computer science. In both fields, the origins of PbD go back long before 2000 [6,9].

4.1 Article 25

PbD as a policy discourse culminated in Article 25 GDPR bearing the title "data protection by design and by default". Article 25 is a general obligation to integrate core data protection principles into the design of data processing architectures. It is seen as one of GDPR's 'most innovative and ambitious norms' [4]. Article 25(1) - the data protection by design (DPbD) part - reads as follows:

> Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to

> implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects.

The range of Article 25 is broad, its language is vague, and it offers only little guidance. Thus it is not straightforward how MPC should be seen in the context of Article 25. To answer this question, we look at it from the privacy engineering (Section 4.2) side and the latest European Data Protection Board (EDPB) guidelines (Section 4.3).

4.2 Privacy Engineering

The task of privacy engineering is to translate Article 25 objectives into concrete design strategies. In their groundbreaking work, Spiekermann and Cranor [33] came up with a framework for designing privacy-friendly systems. Their framework distinguishes between a privacy-by-policy and a privacy-by-architecture approach. The privacy-by-policy approach concentrates on the realization of proper notice, choice, and purpose limitation principles. In contrast, the privacy-by-architecture approach focuses on the implementation of data minimizing by means of reducing identifiability and linkability. Hoepman [25] built upon this framework and derived eight privacy design strategies, four for each approach. These strategies were later adopted by European Union Agency for Cybersecurity's report "Privacy and Data Protection by Design - from policy to engineering" [12]. For our purpose, the four data-oriented strategies (privacy-by-architecture) are of particular interest: minimize, hide, separate and aggregate. To all those strategies, MPC can positively contribute but most naturally to data minimization.

Güres, Troncoso, and Diaz, in their seminal work "Engineering Privacy by Design" [23] and a follow-up article [24], give a blueprint on how to apply data minimization strategies following four activities. The first activity when designing privacy-friendly systems is to classify the system in a user and a service domain. This distinction is essential because, as pointed out in their second article, "data minimization" is an ambiguous term. We often do not aim for data minimization in the information-theory sense. Instead, we want to minimize the amount of personal data that a user has to disclose in order to use a service. In other words, the overall goal is to reduce the need for trust. The next activity is to identify the necessary data for achieving the purpose of the system. Afterward, one has to map the data to the user and service domain. The last activity's goal is to remove as much data as possible from the service domain by using privacy-enhancing technologies.

Privacy-Enhancing Technologies. The European Commission [17] defines PETs as technology that 'can help to design information and communication systems and services in a way that minimises the collection and use of personal data and facilitate compliance with data protection rules.' Similar to privacy

design strategies, the diverse set of PETs can roughly be grouped into two categories [14]. Soft PETs aim to enforce data subject rights (e.g., transparency, erasure, information) which match with privacy-by-policy objectives. Hard PETs, in contrast, try to minimize personal data in the above sense. From a threat model perspective, we could say that while hard PETs include controllers and processors, soft PETs do not. Since MPC places limited trust in controllers and processors, it is a prime example of a hard PET.

Rubinstein and Good [30] argue that Article 25 obligates controllers to adopt not only soft PETs but also hard PETs, assuming they are both available and suitable for the task at hand. Their argumentation can be summarized as follows. Article 25 requires that controllers shall implement 'appropriate technical and organizational measures' that 'implement data-protection principles, such as data minimization, in an effective manner'. As pointed out by the EDPB, 'effectiveness is at the heart of the concept data protection by design' [18]. Hard PETs are more effective since they offer strong technical guarantees, whereas soft PETs can only offer vague policy commitments.

4.3 EDPB Guidelines

We now summarize the most relevant points of the recent EDPB guidelines on "Data Protection by Design and by Default" [18] concerning MPC. The most compelling argument against MPC is its computational overhead over plain[7] solutions and, subsequently, its higher monetary costs. While Article 25 allows the controller to take into account 'the cost of implementation', the EDPB clarifies that costs can only guide the decision on how to implement DPbD. Not if controllers and processors should implement DPbD at all. Consequently, the cost of MPC has to be compared to alternatives providing the same effectiveness on data minimization rather than the plain implementation. Another often neglected fact is that 'DPbD applies at existing systems that are processing personal data', and this even includes 'systems pre-existing before the GDPR entered into force'[18]. Lastly, if GDPR breaches occur, DPbD has an impact on the level of monetary sanctions.

4.4 Conclusion

We have seen strong evidence that MPC is a perfect match to satisfy the obligations resulting from Article 25 (1) - DPbD. First of MPC is an effective PET for the purpose of data minimization and can be considered state-of-the-art. Moreover, the EDPB guidelines provide very compelling arguments for an MPC adoption. This is true especially for public administrators as the EDPB reemphasizes Recital 78 that public administrators should lead by example.

[7] plain solution refers to a solution without MPC.

5 MPC in the GDPR

This section aims to devise a test to determine the impact of an MPC protocol with regard to the GDPR.

5.1 Related Work

Most closely related to this work is an article by Spindler et al. [34], who analyzed the role of personal data and encryption in the GDPR. A subsection is devoted to MPC. They conclude that the GDPR is not applicable to MPC in the relative approach. Also, our analysis comes to the same conclusion for a specific class of use cases. However, there are many use cases where we believe that the GDPR applies to MPC. This divergence, in opinions, could be based on Spindler et al.'s neglection to acknowledge that MPC only protects data during the computation but not the computation's output. We show that the output of an MPC could still be personal data, even in the relative approach.

Our paper shares a similar objective with work by Nissim et al. [29]. They manage to bridge the gap between legal privacy requirements and a mathematical privacy model. In particular, they show that a specific PET - differential privacy [35] - satisfies the privacy protection set forth by the US regulation Family Educational Rights and Privacy Act 1974 (FERPA) [1]. A follow-up work [7] concluded that differential privacy most likely satisfies "singling out" which is a concept in the GDPR. Nevertheless, we do not expect a similar result for the GDPR as for FERPA. In our context, both MPC as PET and the GDPR as legislation are way broader than differential privacy and FERPA. To sum up, a use case independent result seems rather unrealistic.

5.2 Test

In order for the test to be concrete and concise, there have to be a few assumptions. The test is designed from the (potential) controller or processor point of view, i.e., should the MPC protocol involve multiple controllers or processors, each of them has to go through the test. Further, the test assumes that the MPC protocol is secure. What exactly is meant by secure is discussed in Section 5.3. The following test is depicted in Figure 2.

Absolute vs. Relative. The first question is which approach to follow. In the absolute approach, the GDPR applies as soon as one party inputs personal data. It does not matter if the personal data is in a secret shared form or any other technique to realize MPC protocols. Because the data is personal data for at least one party, it automatically is for all other parties, assuming the absolute approach. If no personal data is involved, the computation is out of the GDPR's scope regardless of the use of MPC. In contrast, the situation is more subtle in the relative approach. Namely, here the specifics matter who provides which input and, even more importantly, who receives what kind of output.

Input Data. In the relative approach, the first question a controller or processor should ask herself is about the nature of the input data. It makes a difference if the input data provided constitutes personal data. Notwithstanding that MPC will protect the input data during the computation, one must be careful about the output. It could result in a transfer of personal data.

Output Data. The most crucial part of the assessment represents the questions regarding the output. Recall that the output of every meaningful computation, hence including MPC, discloses some information.

There are up to three questions a controller or processor has to answer. The first two are only relevant in the relative approach. If the controller or processor provides personal data, it matters greatly if any other party receives personal data via the output. In such a case, personal data is transferred from the controller or processor to a third party. Thus the GDPR applies to the entity performing the test as well as the recipient of the personal data. The relationship under the GDPR (e.g., joint controllers) between involved parties can now be assessed independently of this test.

The second question determines whether the controller receives personal data from a third party via the output. It must only be considered if the controller or processor does not input personal data or nobody else receives personal data. Arriving at this question and answering it with no will lead to the non-applicability of the GDPR for this MPC protocol. The GDPR can be avoided because no personal data is transferred to another party or received from another party, and MPC protects potential personal data input during the computation.

Data Minimization. The remaining question tries to answer whether the particular MPC protocol implements DPbD. If, until now, the applicability of the GDPR could not be ruled out, this question is relevant in the relative and absolute approach. The controller or processor has to check if the personal data output to the parties (including herself) is minimized through the use of MPC. Only in the affirmative case is MPC a suitable candidate for DPbD. Note that this minimization should not be seen strictly in the information-theoretic sense. Instead, it aims to ensure that MPC lowers the risk to the rights and freedoms of the data subject(s).

5.3 Security Model and Trust Assumption

Which security model and trust assumption are sufficient depends on the use case. From a data protection point of view, a very rigorous approach is always preferable. More specifically, the most reliable data protection is given if the MPC protocol is maliciously secure. In addition, the trust assumption should be reduced as far as possible to a level where the protocol is secure as long as one party does not collude with the others.

MPC protocols fulfilling the properties above are highly complex and involve significant computational overhead. Thus, an interesting question is if there are

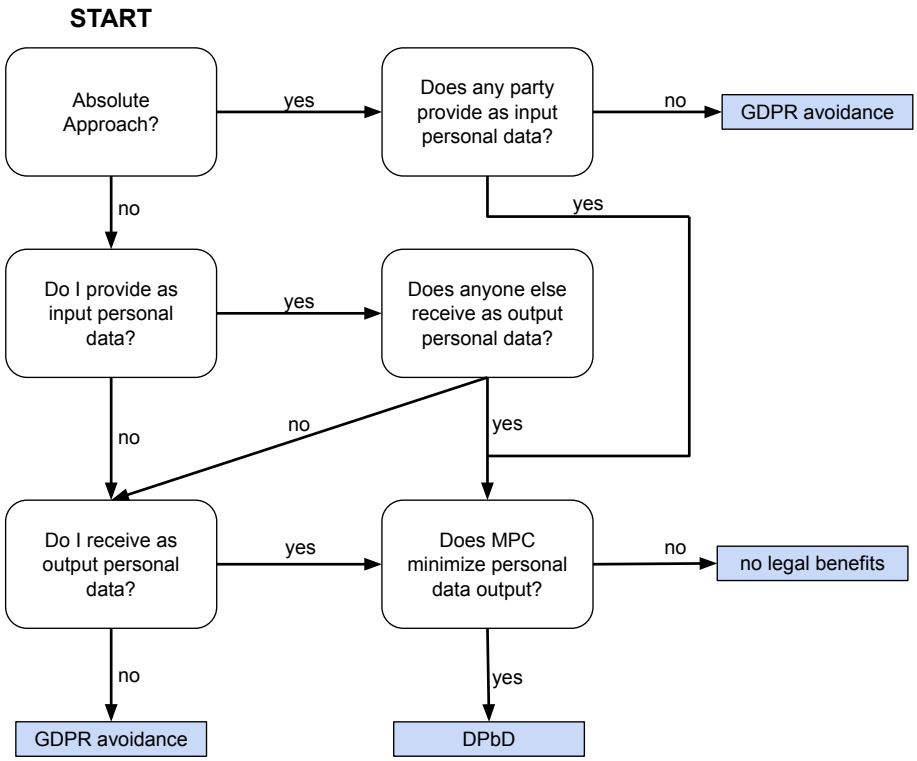

Fig. 2. Assessment scheme for MPC in the GDPR

use cases where semi-honest MPC protocols or a different trust assumption suffices. The following paragraphs should by no means be seen as the last word on the subject. Rather, it is intended to spur an interdisciplinary discussion. In our view, it cannot be solved from a purely technical point of view as legal and normative considerations have also to be taken into account.

GDPR Avoidance. If MPC is applied to avoid the GDPR, we should demand exceptional robust security guarantees. Because if a security violation happens, personal data may be exposed. In addition, since initially, the GDPR was not applicable for this computation, there is the possibility that neither party has any procedures in place to mitigate such exposure of personal data (notification and communication of personal data breach - Article 33, 34 GDPR). Thus, we recommend that only maliciously secure protocols can lead to the non-applicability of the GDPR. Also, we would prefer that colluding parties cannot compromise security. However, at least the parties should be legally bound by law or contract not to collude.

Data Protection by Design. If MPC is applied for DPbD objectives, we can discuss less strict requirements. Here the use of MPC is to achieve data minimization. Controllers and processors should be encouraged to deploy MPC protocols and not be scared by too high standards. Hence, we advocate that in such a case, even semi-honest secure MPC protocols could suffice. Further, the trust assumption could also be relaxed. For instance, one could assume that parties do not collude if it is diametral to their business interests. This fact should be checked regularly as it is subject to change.

6 Scenarios

In this section, we present two common scenarios where MPC is of particular interest. Both use cases involve only two parties to make the analysis easier to follow. Nevertheless, there would be no substantial difference in applying our test for more than two parties.

6.1 Private Set Intersection

Two public authorities, A and B, perform joint data analysis. Authority A has a database containing records about vaccinated individuals. On the other side, authority B holds a database consisting of individuals who have diabetes (see Figure 3 for a small example). The goal of the data analysis is that authority B can answer the following question.

What is the vaccination rate among individuals who have diabetes?

Solution 1. Authority A and B engage in a private set intersection protocol. Similar to the protocol in see Section 2.2, the only difference is that now just one party - authority B - receives the intersection. So the protocol's output will be the identifiers (e.g., names) of individuals that are vaccinated and have diabetes. Thus authority B can compute the percentage of vaccinated rate among people with diabetes.

Analysis: To determine the solution's impact with regard to the GDPR, we follow the test from above. We start from the perspective of authority B and assume the relative approach. Clearly, the information on whether someone has diabetes constitutes personal data. Authority A does not receive an output from the computation. However, authority B itself ends up knowing the names of the intersection (the names underlined in Figure 3). It can then deduce that every individual in that intersection is vaccinated.

So, we arrive - as we would also have in the absolute approach - at the question of whether the MPC solution minimizes the personal data output. It does indeed because, in the naive solution, authority A would send the complete database to authority B. Consequently, authority B would know not only who

Authority A

Vaccinated
Maynard Collins
Breanna Sanders
Buster Joyce
Devon Robertson
Georgia Judd
Napoleon Blakely
Callan Fitzroy
Bryon Morin

Authority B

Diabetes
Leeann Marsden
Breanna Sanders
Seymour Barlow
Devon Robertson
Darby Samson
Napoleon Blakely
Glen Hoggard
Bryon Morin

Fig. 3. Medical Conditions Databases

from its diabetes database is vaccinated but who is vaccinated in general. Thus, the use of this MPC solution qualifies as DPbD. The outcome for authority A is the same, albeit a slightly different argumentation. Namely, authority A does not receive personal data but transfers personal data (vaccination statuses) to authority B.

Solution 2. It is based on the previous solution, with the difference being that the output will be the size of the intersection. In other words, authority B only learns the number of vaccinated individuals in its diabetes database but not their names (50% in our small example).

Analysis: In the absolute approach, the new solution does not change the outcome of the above analysis. The more interesting case is when we consider the relative approach. The crucial difference compared to the previous solution is that the protocol's output is now non-personal data, provided that the databases are not artificially small. Hence, neither authority A nor authority B receives personal data through the computation. Thus, this computation does not deal with personal data at all. Therefore it falls outside of the scope of the GDPR.

Solution 3. One should not get the impression from this example that MPC is always favorable. To show this, we construct a hypothetical MPC solution to the problem above. It is important to mention that no serious privacy engineer would propose such a solution. One could design an MPC protocol that outputs A's database to B.

Analysis: The computation would still be done by means of MPC, but that protection would be void since the output reveals all the personal data anyhow. Obviously, in this case, MPC does not minimize the personal data output in any form. Thus, although a PET was used, there should be no legal benefit from it.

6.2 Outsourcing

Jane Doe is concerned about one of her moles. She takes a photo of it with her cell phone. Afterward, she uploads it to the cloud service MoleChecker. Based on their classification algorithm, they tell her how likely her mole is cancerous.

Solution. One can design an MPC protocol with the following properties. The protocol's output is still the likelihood of the mole being cancerous, but now the output is only received by Jane Doe. So MoleChecker does not get the picture (protected by MPC) nor the result.

Analysis: Jane Doe is a data subject, and therefore we solely check the perspective of MoleChecker in its role as a potential controller. The photo is personal data for Jane Done. Consequently, it is also for MoleChecker, assuming the absolute approach. Hence, the GDPR applies to this computation, and MoleChecker becomes a controller. Since the MPC solution minimizes personal data output (MoleChecker never sees the photo or learns the result), it qualifies as DPbD.

If we change to the relative approach, another picture emerges. Because MoleChecker does not input personal data or receive any through the MPC protocol, the GDPR does not apply in this situation (accordingly, MoleChecker is not a controller).

Variant. Finally, we take a look at a variant of this image classification use case. The functionality stays the same, but the setting is slightly different. A hospital outsources the classification of MR images to a company called MRClassifier. Since the solution is identical from the technical perspective, the analysis is as well. However, it is interesting what this means for the involved parties. In the absolute approach, the hospital becomes a controller, which turns the MRClassifier into a processor. Accordingly, a data processing agreement between the parties is mandatory. Since, as already seen above, the GDPR does not apply to this computation in the relative approach, no such agreement is necessary.

7 Conclusion

We believe the use of MPC would lead to better data protection without significant restriction on data-driven business opportunities. Hopefully, this paper contributes to more legal certainty if applying MPC. To see more widespread adoption of MPC, we can only back EDPB's recommendation for certification [18]. The certifying of MPC protocols has two benefits. It would be a guidance for controllers on how to use MPC properly in their processing operations. Secondly, data subjects could better see for themselves which controllers follow best-practice data protection measurements. At best, our work starts a discussion on MPC certification and standardization with the goal of better data protection for individuals. In the end, such an initiative can only be successful if it is highly interdisciplinary, involving MPC, legal, privacy engineering, and domain experts as well as the data protection authorities.

In a future line of work, we would like to investigate the role of MPC in the upcoming e-privacy regulation.

Acknowledgments

We thank the reviewers of the Privacy Symposium for their comments, which helped improve the paper's quality. We also thank Aisling Connolly for many fruitful discussions in the early stages of this work.

This work was supported by EU's Horizon 2020 project Safe-DEED, grant agreement n°825225 and TRUSTS grant agreement n°871481.

References

1. Family educational rights and privacy act of 1974, 20 u.s.c.§1232g (2012)
2. Article 29 Working Party: Opinion 05/2014 on anonymisation techniques (2014), https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf
3. Bergauer, C., Gosch, N.: Die pseudonymisierung personenbezogener daten gemäß der dsgvo (2020)
4. Bygrave, L.A.: Data protection by design and by default: Deciphering the eu's legislative requirements. Oslo Law Review **4**(02), 105–120 (2017)
5. Cavoukian, A., et al.: Privacy by design: The 7 foundational principles. Information and privacy commissioner of Ontario, Canada **5**, 12 (2009)
6. Chaum, D.: Security without identification: Transactiooffersn systems to make big brother obsolete. Communications of the ACM **28**(10), 1030–1044 (1985)
7. Cohen, A., Nissim, K.: Towards formalizing the gdpr's notion of singling out. Proceedings of the National Academy of Sciences **117**(15), 8344–8352 (2020)
8. Commission, D.P.: Guidance on anonymisation and pseudonymisation (2019), https://www.dataprotection.ie/sites/default/files/uploads/2019-06/190614%20Anonymisation%20and%20Pseudonymisation.pdf
9. Council of European Union: Council of europe convention 108: Convention for the protection of individuals with regard to automatic processing of personal data 1981, ets 108 (1981), https://www.coe.int/en/web/conventions/full-list/-/conventions/rms/0900001680078b37?module=treaty-detail&treatynum=108

10. Cramer, R., Damgård, I.B., et al.: Secure multiparty computation. Cambridge University Press (2015)
11. Daemen, J., Rijmen, V.: Aes proposal: Rijndael (1999)
12. Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J.H., Metayer, D.L., Tirtea, R., Schiffner, S.: Privacy and data protection by design-from policy to engineering. arXiv preprint arXiv:1501.03726 (2015)
13. De Cristofaro, E., Tsudik, G.: Practical private set intersection protocols with linear complexity. In: International Conference on Financial Cryptography and Data Security. pp. 143–159. Springer (2010)
14. Deng, M., Wuyts, K., Scandariato, R., Preneel, B., Joosen, W.: A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. Requirements Engineering **16**(1), 3–32 (2011)
15. Dwork, C.: Differential privacy. In: International Colloquium on Automata, Languages, and Programming. pp. 1–12. Springer (2006)
16. European Commision: Eu study on the legal analysis of a single market for the information society (2014), `https://op.europa.eu/s/sA5L`
17. European Commission: Promoting data protection by privacy enhancing technologies (pets) (2007), `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52007DC0228`
18. European Data Protection Board: Guidelines 4/2019 on article 25 data protection by design and by default (2020), `https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-42019-article-25-data-protection-design-and_en`
19. European Parliament and of the Council: Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation), oj 2016 l 119/1 (2016), `https://eur-lex.europa.eu/eli/reg/2016/679/oj`
20. Evans, D., Kolesnikov, V., Rosulek, M.: A pragmatic introduction to secure multiparty computation. Foundations and Trends® in Privacy and Security **2**(2-3) (2017)
21. Finck, M., Pallas, F.: They who must not be identified—distinguishing personal from non-personal data under the gdpr. International Data Privacy Law (2020)
22. Goldwasser, S., Micali, S., Rackoff, C.: The knowledge complexity of interactive proof systems. SIAM Journal on computing **18**(1), 186–208 (1989)
23. Gürses, S., Troncoso, C., Diaz, C.: Engineering privacy by design. Computers, Privacy & Data Protection **14**(3), 25 (2011)
24. Gürses, S., Troncoso, C., Diaz, C.: Engineering privacy by design reloaded. In: Amsterdam Privacy Conference. pp. 1–21 (2015)
25. Hoepman, J.H.: Privacy design strategies. In: IFIP International Information Security Conference. pp. 446–459. Springer (2014)
26. Kamm, L., Bogdanov, D., Laur, S., Vilo, J.: A new way to protect privacy in large-scale genome-wide association studies. Bioinformatics **29**(7), 886–893 (2013)
27. Mohassel, P., Zhang, Y.: Secureml: A system for scalable privacy-preserving machine learning. In: 2017 IEEE symposium on security and privacy (SP). pp. 19–38. IEEE (2017)
28. Mourby, M., Mackey, E., Elliot, M., Gowans, H., Wallace, S.E., Bell, J., Smith, H., Aidinlis, S., Kaye, J.: Are 'pseudonymised'data always personal data? implications of the gdpr for administrative data research in the uk. Computer Law & Security Review **34**(2), 222–233 (2018)

29. Nissim, K., Bembenek, A., Wood, A., Bun, M., Gaboardi, M., Gasser, U., O'Brien, D.R., Steinke, T., Vadhan, S.: Bridging the gap between computer science and legal approaches to privacy. Harv. JL & Tech. **31**, 687 (2017)
30. Rubinstein, I.S., Good, N.: The trouble with article 25 (and how to fix it): the future of data protection by design and default. International Data Privacy Law (2020)
31. Sangers, A., van Heesch, M., Attema, T., Veugen, T., Wiggerman, M., Veldsink, J., Bloemen, O., Worm, D.: Secure multiparty pagerank algorithm for collaborative fraud detection. In: International Conference on Financial Cryptography and Data Security. pp. 605–623. Springer (2019)
32. Smart, N.P., Smart, N.P.: Cryptography made simple. Springer (2016)
33. Spiekermann, S., Cranor, L.F.: Engineering privacy. IEEE Transactions on software engineering **35**(1), 67–82 (2008)
34. Spindler, G., Schmechel, P.: Personal data and encryption in the european general data protection regulation. J. Intell. Prop. Info. Tech. & Elec. Com. L. **7**, 163 (2016)
35. Wood, A., Altman, M., Bembenek, A., Bun, M., Gaboardi, M., Honaker, J., Nissim, K., O'Brien, D.R., Steinke, T., Vadhan, S.: Differential privacy: A primer for a non-technical audience. Vand. J. Ent. & Tech. L. **21**, 209 (2018)
36. Yao, A.C.: Theory and application of trapdoor functions. In: 23rd Annual Symposium on Foundations of Computer Science (SFCS 1982). pp. 80–91. IEEE (1982)