Wouter Seinen and Ian Wachters – Pinsent Masons Netherlands and Roseman Labs

Overcoming data sharing & transfer risks with Multi Party Computing

Data collaboration and PETs[1]

In today's world, data is generated in almost every activity we undertake. This data often resides in silos of different organizations, sometimes across jurisdictions. McKinsey estimated that if we were to share and collaborate on this data, $3.000 billion in value could be unlocked. A lot of that value also includes societal benefits, such as better healthcare and more effective crime fighting.

How tempting generating positive impact might be, data sharing initiatives often strand on the cliffs of regulatory compliance. Legislations like GDPR as amplified by court decisions as in the Schrems II case made organizations realize that straightforward data sharing is often very complex to arrange in a compliant manner. And when parties in third countries also involved, sharing is generally considered prohibited.

Equally problematic are use cases of sharing data with organizations in countries with data localization requirements. Russia and China are most known for their national rules that require data being stored on local servers, but there are several other countries who also require certain data types to be stored locally.

Third, one of the most common data transfer scenario's, since the Schrems II decision and the EDPB recommendations on supplemental transfer tools have confirmed, is the use of shared IT systems within international groups of companies.

Certain Privacy Enhancing Technologies (PETs) enable parties, owning different sets of sensitive data, to collaborate on this data in a privacy-preserving and secure manner. They can generate insights without actually sharing their data with each other.

These PETs are based on cryptography and have been developed by academia over the last 40 years. Now, they have become sufficiently mature and affordable for the industry to commercialize their application at scale.

Sharing insights without transferring sensitive data

Roseman Labs, a Dutch start-up company, has developed a solution based on Multi Party Computation (MPC). MPC is one of the strongest and most performant PETs. It combines strong technical measures (processing of encrypted data) with hard-coded segregation of duties. With MPC, data is fragmented into so-called 'secret shares', which reside in multiple MPC servers that can jointly execute computations without centralizing the data in a single location. Secret shares are random data, not disclosing anything about the source data. The magic of MPC is that the servers can perform joint calculations on these secret shares, without revealing the source data at any moment in time.

One recently implemented use case involves a public-private partnership between NGOs fighting human trafficking (Sustainable Rescue and other parties that prefer to stay unnamed) and a dedicated team of the Dutch National Police. All parties have data on (potential) victims of human trafficking, yet some of these are informants of the NGOs. The police would want to know who the informants of the NGOs are, as this would allow them to not follow these individuals but focus their resources on others. However, sharing the names of the informants is impossible as both the police and the NGOs must abide their duty of confidentiality. With and on behalf of the public-private partnership, Roseman Labs developed a solution to enable comparison of names under encryption, thereby ensuring that those victims selected by the police are not active informants of the NGOs.

Other applications of Multi Party Computation

MPC provides a powerful way to collaborate on data in a privacy-preserving manner. It enables organizations to collaborate on data even when it cannot, may not or will not be shared otherwise. Such sensitive data processing could relate to, for example, anti-money laundering, anti-trust laws, or data that is not allowed to leave a certain jurisdiction (e.g. due to data localization laws). Example use cases are plentiful, and rapidly emerging in real life:

- Healthcare: privacy-sensitive patient data resides with different care providers. To develop a better understanding of treatment effectiveness, data are typically replicated in dedicated (costly) studies. The movement of such real-world evidence aims to reuse existing patient record data. MPC enables the use of this real-world data while protecting patient privacy.
- Law enforcement: Information about criminal activities and fraud is scattered across law enforcement agencies, banks and private businesses. Combining this informa-

[1] We thank Mr. Andre Walter, Roderick Rodenburg and Paul Fockens, who were all involved in the case study elaborated on in this article, for their contributions.

tion to spot patterns and avoid repeat actions by the same criminal is very difficult today. Again, MPC enables different organizations to combine their information, without disclosing unnecessary details.
- Cyber security: Cyber threat intel is highly sensitive. No organization will disclose details of how they have been attacked. The Dutch National Cyber Security Center (NCSC) is deploying MPC to collect and consolidate cyber threat intel information from more than 100 (growing to 15.000) organizations and businesses in the Netherlands in a confidential manner. With this, the NCSC can distill trends and inform organizations almost instantly about active threats.
- Financial services: Under intense regulatory pressure, banks have expanded their financial crime detection teams to thousands of employees. Unfortunately, these teams largely operate in silos. Smart criminals exploit this by using multiple banks to launder their money. MPC, in this case, offers a secure way for financial institutions to work together more closely against money laundering.
- International data transfers: Often, data is transferred because there is a need for central reporting. However, in many use cases data does not have to be accessed on record-level. In other words, the analyses and reports that are run on a global database often do not contain personal data; such data was only used to make the computations to generate the report at hand or run the statistical analysis. When deploying MPC technology, the central database can be split in local databases, so that each organization will only have 'their own' dataset, whilst over-arching reports and analyses can still be done. Since the European Data Protection Board has specifically mentioned MPC as a suitable technical measure to avoid transfer of personal data to problematic jurisdictions, organizations will have to consider the viability of this technology in their data transfer impact assessments.

The legal perspective – what's different with MPC?

What differentiates MPC? Why is it that organizations can suddenly collaborate on data with MPC, while this was not possible before? How does this work from a legal perspective?

It is important to understand that, when using MPC and processing data on secret shares, data is still considered personal data. The secret shares cannot be considered anonymous because the data can be reconstructed to its original form if the majority of all data owners collude. They remain personal data and so privacy regulations such as the GDPR still apply.

That being said, compliance with GDPR regulations becomes easier and more robust. Below we provide examples of typical requirements for dealing with personal data and how they can be enforced through the use of MPC:
- Purpose binding: Whenever personal data is processed, it needs to be clear what the purpose of the processing is. In traditional approaches, it is almost impossible to control what the data is used for once a copy of the data is shared - whether it is for the given purpose or beyond. With MPC, the data itself is not shared. Only a specific calculation, approved by the data owner, is allowed on the data. This is a very strong way of enforcing purpose binding.
- Data control: MPC does not require multiple copies of the data. Processing is performed directly on the source data. If the source data changes, the operation immediately includes that change. If a data owner decides to withdraw from the cooperation, they simply stop approving calculations on their data.
- Data minimization and proportionality: Only the results of the analysis are shared, not the underlying data. This leads to a vast reduction of data exposure. It makes the use of data far more proportional compared to a situation where all data must be exposed in order to get the same result.
- Data localization: Data in its original form does not leave the technical environment of the data owner. It is made available in the secret shares and combined with other parties' data while keeping their own data locally. In other words, it is the calculation on that data that travels, not the data itself. Often, the results are not personal data anymore (because of aggregation) and can be shared.
- Technical measures: MPC is a strong form of encryption that even holds in a post-quantum world. It can be proven mathematically that data remains secret, as long as a majority of the MPC servers do not collude.
- Organizational measures: The strong technical features of MPC can be further strengthened with the right

> "Few legal, compliance and data protection officers are familiar with MPC."

organizational measures. Segregation of server access across the organizations of the data controllers results in a system in which none of the users are able to decrypt the data on their own. Data owners naturally want to protect their own data, so there is no incentive for them to collude with the other parties. Furthermore, data access and approval of the analysis are hardwired into a process of collecting digital signatures from all data owners. Without these signatures, no analysis can be performed on the data.

Today, few legal, compliance and data protection officers are familiar with MPC, and few tech experts oversee all legal implications. To reap the full benefits of this new technology, knowledge sharing is necessary.

> “With modern techniques like MPC, new insights can be gained.”

Conclusion

Initiatives to generate value from shared data sets for legitimate purposes often strand because of GDPR compliance concerns. With modern techniques like MPC, new insights can be gained while protecting the source data and its subjects. MPC technology is getting more mature and operational solutions are being implemented right now.

Today, the most obvious fields of applications include: (1) sharing sensitive data beyond what is currently possible, (2) international data transfer of personal data to and from problematic jurisdictions and (3) addressing data-localization restrictions imposed by third country laws.

We believe that the field of application of MPC will evolve far beyond discrete use cases like the case study above. Amongst the most common sharing scenarios are intra-group data sharing for management purposes and the use of so-called blacklists by various organizations that are unaffiliated. There is much controversy around the use of blacklists: legal obligations to investigate and report suspicious behavior and screen customers are mushrooming, whilst on the other hand there is resistance against the use of central systems with this type of information. The Dutch DPA has granted some licenses for blacklists, but the general rule is that cross-sector exchange of blacklist data is prohibited. Using MPC to create decentralized blacklists may be a solution to reduce unnecessary data sharing without defeating the legitimate purpose behind the blacklist.

There are many other (international) data sharing use cases where MPC could potentially resolve very thorny compliance concerns. We hope that the legal and tech community will engage on the topic and step up their collaboration to further accelerate the adoption of MPC and other privacy enhancing technologies.

About the author
Wouter Seinen is a technology lawyer at Pinsent Masons Netherlands. Wouter heads the Amsterdam office and leads the Cyber, Data and Technology practice in the Netherlands. He advises data-driven business on a broad range of matters involving technology and law, with a focus on commercial partnerships, data monetization and data protection.

About the author
Ian Wachters is Commercial Officer at Roseman Labs. After a corporate career of more than 25 years (Shell and Boston Consulting Group), Ian now works with ambitious tech-driven start-ups that aim to make the world a better place. In his last role at BCG he built and managed BCG’s pricing activities across EMEA and Latam with $100M in revenues.