# **Homomorphic Encryption**

Homomorphic encryption is a special type of encryption that supports computation on encrypted data (ciphertexts) without decryption. Thanks to this property, homomorphically encrypted data can be securely handed out to third parties, who can perform meaningful operations on them without learning anything about their content. Fully homomorphic encryption schemes, or schemes enabling arbitrary computations on ciphertexts, are still considered nonviable due to the high computational and storage overheads they introduce. Current practical schemes that enable only a limited number of computations on ciphertexts (such as polynomial operations) have reached a level of maturity that permits their use in real scenarios.

# **Secure Multiparty Computation**

Secure multiparty computation protocols enable multiple parties to jointly compute functions over their private inputs without disclosing to the other parties more information about each other’s inputs than what can be inferred from the output of the computation. This class of protocols is particularly attractive in privacy-preserving distributed analytic platforms due to the great variety of secure computations they enable. However, this flexibility includes several drawbacks that hinder their adoption, including high network overhead and the requirement for parties to remain online during computation.

# **Multiparty Homomorphic Encryption**

The combination of secure multiparty computation and homomorphic encryption was proposed to overcome their respective overheads and technical limitations; we refer to it as multiparty homomorphic encryption. Multiparty homomorphic encryption enables flexible secure processing by efficiently transitioning between encrypted local computation, performed with homomorphic encryption, and interactive protocols (secure multiparty computation). It can be used to choose the most efficient approach for each step within a given workflow, leveraging the properties of one technique to avoid the bottlenecks of the other. Moreover, multiparty homomorphic encryption ensures that the secret key of the underlying homomorphic encryption scheme never exists in full. Instead, it distributes the control over the decryption process across all participating sites, each one holding a fragment of the key. All participating sites have to agree to enable the decryption of any piece of data, and no single entity alone can decrypt the data.

Unlike homomorphic encryption or secure multiparty computation alone, multiparty homomorphic encryption provides effective, scalable, and practical solutions for addressing the privacy-preserving issues that affect the distributed or federated approach for data sharing. For example, systems such as Helen, MedCo, or POSEIDON use multiparty homomorphic encryption to guarantee that all the information interchanged between the sites is always in encrypted form, including aggregate data such as model parameters and model updates, and only the final result (the computed model or the predictions based on this model) is revealed to the authorized user. Finally, multiparty homomorphic encryption reduces the need of obfuscation techniques to protect aggregate-level data from inference attacks. Furthermore, data utility, which is typically lost with privacy-preserving distributed approaches that only rely upon obfuscation techniques, can be significantly improved. As aggregate-level data transfer and processing across participating sites during the analysis or training phase remains always encrypted, obfuscation can be applied only to the decrypted final result of the analysis that is released to the data analyst, instead of being applied to all local model updates at each iteration. Hence, multiparty homomorphic encryption enables a much lower utility degradation for the same level of reidentification risk.

# **Trusted execution environments**

A trusted execution environment (TEE) is a secure area inside a main processor. TEEs are isolated from the rest of the system, so that the operating system cannot read the code in the TEE. However, TEEs can access memory outside. TEEs can also protect data “at rest,” when it is not being analyzed, through encryption.

Like homomorphic encryption, TEEs might be used to securely outsource computations on sensitive data to the cloud. Instead of a cryptographic solution, TEEs offer a hardware-based way to ensure data and code cannot be learnt by a server to which computation is outsourced. Unlike homomorphic encryption, current TEEs are widespread and permit the computation of virtually any operations.

Commercial organizations, for example, face this tension when there is value in learning from the practices and performance of competitors but accessing relevant data would reveal commercially sensitive information. Certain cryptographic techniques can enable different organizations to analyze and derive insights from data without pooling it or sharing it with each other. This opens up the potential for companies to learn from each other without giving away trade secrets.

**Technologies suitable for health information sharing**

## **Data Anonymization and Pseudonymization**

The GDPR defines *personal data* as concerning an identifiable natural person. Therefore, pseudonymized data, where all identifiers have been removed from those data, remain personal data. However, the provisions of the GDPR do not concern anonymized data or data which have been processed so individuals are no longer identifiable. In particular, anonymized data may be used for research or statistical processing without the need to comply with the GDPR.

## **Data Processing**

The GDPR’s provisions apply to data controllers, or entities determining the purpose and means of processing personal data. This definition encompasses both health care institutions and research institutions. Data controllers must guarantee personal data processing is lawful, proportionate, and protects the rights of data subjects. In particular, the GDPR provides that encryption should be used as a safeguard when personal data are processed for a purpose other than which they were collected. Although the GDPR does not define encryption, the Article 29 Working Party treats encryption as equivalent to stripping identifiers from personal data. The GDPR also lists encryption as a strategy that can guarantee personal data security. Furthermore, the GDPR emphasizes that data controllers should consider the state of the art, along with the risks associated with processing, when adopting security measures. The GDPR also provides that data processing for scientific purposes should follow the principle of data minimization. This principle requires data processors and controllers to use nonpersonal data unless the research can only be completed with personal data. If personal data are required to complete the research, pseudonymized or aggregate data should be used instead of directly identifying data.

## **Advanced Privacy-Enhancing Technologies and EU Data Governance Requirements**

In this section, we argue that multiparty homomorphic encryption, or homomorphic encryption and secure multiparty computation used in concert, meets the requirements for anonymization of data under the GDPR. Furthermore, we argue the use of multiparty homomorphic encryption can significantly reduce the need for custom contracts to govern data sharing between institutions. We focus on genetic and clinical data sharing due to the potential for national derogations pertaining to the processing of health-related data. Nevertheless, our conclusions regarding the technical and legal requirements for data sharing using multiparty homomorphic encryption, or homomorphic encryption and secure multiparty computation, may apply to other sectors, depending on regulatory requirements.

**Technologies suitable for governance**

Navigating the tensions created by new uses of data, set out above and in the *Data Management and Use: Governance in the 21st Century* report, requires appropriate governance mechanisms, from codes of conduct and ethics to regulation. However, in some cases, technological solutions can help diffuse dilemmas between making use of data and protecting both the individuals and organizations that generate or are subjects within datasets. PETs as a category comprises a broad suite of technologies and approaches—from a piece of tape masking a webcam to advanced cryptographic techniques. While some are focused on protecting private communications, the report explored a subset of five PETs identified during the scoping of the project as being particularly promising to enable privacy-aware data collection, analysis, and dissemination of results.

The key question of this paper is whether, according to the current state of development and the trajectory of technological development, we can utilize PETs in addressing social and ethical tensions in data use, and thereby use them as tools for governing the ways that data is used. This will involve consideration both of how these technologies can potentially enable governments and others to unlock the value of data, while also recognizing both contingent and in principle limitations on the role of PETs in ensuring well-governed use of data.

How far are these technologies able to underpin the ways that data use is governed in practice? The field of PETs development is moving quickly, and the Royal Society report captures a moment in time where the technologies are maturing and opportunities to use these technologies are beginning to emerge. It may be that some of the technologies surveyed in our report do not achieve their promise in the near term, or that the costs of adoption prove prohibitive, or that other technologies not explored in depth might leapfrog them. However, there are a number of areas where PETs are already in use which were set out in case studies in the report, with examples summarized here.

There are many examples of current uses of these technologies. For example, TEEs are used in mobile phones to process “touch ID” data. They are also an integral part of secure clouds. The following are some specific examples of where organizations have made use of, or promoted, PETs.

For secure MPC, the first real-world application of Sharemind—which uses MPC—was the analysis of key performance indicators for the Estonian Association of Information Technology and Telecommunications (ITL). The ITL proposed collecting certain financial metrics and analyzing them to gain insights into the state of the sector. The member companies expressed concerns over the confidentiality of the metrics, as they would be handing them out to competitors.

In order to share NHS data securely with multiple teams, while maintaining as much as possible the potential usefulness of the data, NHS Digital have been using a de-identification service employing homomorphic encryption. For security reasons, data is de-identified in different “pseudonymization domains” for each different part of an organization. Within one domain, all data with the same base value is replaced with the same “token” (a nonidentifying value). Across domains, the same base value receives different token. Usually, transferring data between domains requires to remove the encryption for the first domain and replace it with the second domain encryption. However, using consistent “tokenization” and partially homomorphic encryption by Privitar Publisher, it is possible to transform data items between any two domains without revealing the base value, even if they have been de-identified by two instances of the de-identification service using different encryption keys.

This methodology allows the de-identification tool set to be deployed to multiple locations across the NHS and makes any data de-identified by any tool from the de-identification tool set potentially linkable with any other data de-identified by any other tool from the tool set.

In an effort to empower consumers, the UK government promoted midata, a Personal Data Store. Launched in 2011, in partnership with multiple organization, the online portal was designed to provide citizens with access and control over data about them. For example, individuals can access the transactional data for their current account, which they can upload to third party price comparison websites to compare and identify the best value.