Accenture Applied Intelligence

Maximize collaboration through secure data sharing

A detailed look at the Privacy Preserving Computation Techniques ushering in a new era of secure data sharing and enterprise collaboration.

Contents

Privacy Preserving Computation (PPC) techniques are a family of very modern cybersecurity techniques that look at how to represent data in a form that can be shared, analyzed and operated on without exposing the raw information. The following pages dive into some of the primary PPC techniques that are gaining prominence today along with key considerations around how and when to use them.

Technique 1

Trusted Execution Environment (Secure Enclave)

What is it?

A **Trusted Execution Environment**, or Secure Enclave as they are sometimes known, is an environment with special hardware modules that allow for data processing within hardware-provided, encrypted private memory areas directly on the microprocessor chip. This is intended to protect data from attack during computation while the data is in a decrypted state, especially in situations where the data owner is not the only one running processes on the chip, for example in shared hardware set ups like the cloud. In a Secure Enclave, only the owning process has the ability to read or change the data in memory. This is very important in cloud environments where the same hardware can be used by multiple virtual machines owned by multiple users. This is especially relevant in light of the vulnerabilities of this sort like **Spectre and Meltdown** discovered in early 2018 that opened up new possibilities for hardware-based data breaches.

Historically the Trusted Execution Environment was used on a small scale for storing passwords, encryption keys and other small pieces of sensitive data because of size limitations. However, now this capability is available on a larger scale from cloud providers – usually alongside or as part of secure database services – that allow the data in the database to be decrypted only within the Trusted Execution Environment of the servers.

In this way, the data is encrypted at rest and in transit and is kept protected and isolated while unencrypted during computation.

Characteristics

Controlled Environment?

Yes, runs on specialized hardware that limits eavesdropping while the data is decrypted.

Data Obfuscated?

No, all identifying content is still present and accessible to authorized data processors.

Encryption During Processing?

No, data must be decrypted to access but is done inside a tightly controlled space.

What benefits does it provide?

This capability provides protection, especially in a cloud context, to ensure the privacy of data is kept secure during processing.

What are its limitations?

This is suitable for sharing data with highly-trusted parties as all aspects of the shared data are visible to the data recipient. It is only suitable for data sharing scenarios where trust is very high and where the risk of data misuse is low.

What's possible now?

Many cloud vendors are **now providing this capability** as a dedicated low-level service aligned with their computation offerings to try and protect against "in-computation" style attacks in shared hardware. However, there are still vulnerabilities specifically targeting the Trusted Execution Environment (currently theoretical and difficult to execute or exploit) being discovered at the hardware and chip level which are muddying the waters, such as **SGX-ROP** and **SWAPGSAttack**.

Along with the cloud providers, other vendors like **Citrix**™ and **Snowflake**℠ are providing data sharing platform solutions that allow businesses to host their data with the vendor and give fine-grain controls and monitors that allow the business to choose what to share and with whom – all while outsourcing the complexity of managing the enclaves.

Technique 2

Differential Privacy

What is it?

Differential Privacy is a data obfuscation mechanism – often used with other traditional anonymization or de-identification techniques – that allows broad statistical information to be gathered and inferred from data without the actual specifics of individual items being exposed. It does this by introducing additional, fake data or "noise" to the dataset in a very specific way that doesn't change the broad statistical properties of the dataset as a whole. This makes it very difficult to identify individual records from the aggregated dataset. The noise can either be added directly to the data to change each record slightly or added via new synthetic records that artificially increases the total number of records in the dataset.

For example, sharing information about the paths that users take through a set of pages in a website could allow individual user actions to be inferred. The differential privacy model, applied correctly, warrants that even if someone has complete information about 99 of 100 items in a data set, they still cannot reliably deduce the information about the final item. The noise that's added, might, for instance be to add 100 items to the list of fake information that mirrors the behavior of the 100 real items – in this way, statistically, the same percentage of people took one route, but it is very difficult to determine the real items from the "noisy" items that were introduced.

Characteristics

Controlled Environment?

No, the data is treated with the expectation that it will be used outside the control of the data owner.

Data Obfuscated?

Yes, the data is modified so that individual records cannot be identified or de-anonymized.

Encryption During Processing?

No, the data is provided in the clear (unencrypted), but generally anonymized.

Major Variants

The Laplace Mechanism

adds noise to the output of each data record, slightly perturbing the truth in a way that means the value has a very high probability of being very close to the real value.

The Exponential Mechanism

uses a slightly different algorithm which needs more noise to be added to achieve the same level of privacy as the Laplace Mechanism, but it works well where adding noise to particular data fields will render the data unusable (i.e. the weights of a neural net).

What benefits does it provide?

Differential Privacy is about removing certainty of the real data values that make up the dataset when presenting aggregated information and shrouding them within data which is very close to the truth. This means that even if an attacker has information about the data from other sources, they would not be able to correlate it with individual records.

What are its limitations?

It is only useful for statistical processing or data aggregation as the noise added will change the real data but keep its statistical properties. Therefore, it will not work for use cases where the data integrity of specific values within the records is a fundamental part of what needs to be shared.

What’s possible now?

These techniques are being actively used by companies who publish large, open source datasets (i.e. **Apple**, **Google**, **government statistical organizations**) and are being increasingly adopted to avoid privacy leakage and reidentification attacks.

Technique 3

Homomorphic Encryption

What is it?

Homomorphic Encryption enables computation on encrypted data without the need to decrypt it first (or at all). In this way, the sensitive data are encrypted and protected at all stages of transport and processing. The encrypted data can be processed, augmented or changed while still being encrypted by a third party who does not get to see the data they are working with. This mechanism also protects the outputs of the processing as they remain encrypted and are only accessible by the data owner, not the processor. This means that even the most sensitive of data can be shared with a third party without actually exposing it. This mechanism also alleviates the concerns that the third party could learn or derive additional information from the shared data that was not intended.

Homomorphic Encryption works by taking advantage of certain cryptographic properties of encryption algorithms. It allows operations such as one where two encrypted numbers can be added in a way that, when the result is decrypted, would be the same as if the two unencrypted numbers were added. This technique allows a person who can't see the actual data, only the encrypted version of it, to run processes that change the encrypted data without corrupting it. Only low-level operations, like multiplication, addition and subtraction are useable in this fashion, but these base operations can be combined to achieve quite sophisticated results. The specific algorithms that support these techniques often have the further advantage of not currently being considered susceptible to attack from quantum computers.

Characteristics

Controlled Environment?

No, it is specifically designed to act outside the data owner's control.

Data Obfuscated?

Yes, none of the information held in the data is available to the processor.

Encryption During Processing?

Yes, the data does not get decrypted while it is outside the owner's control.

Major Variants

Partially Homomorphic Encryption

schemes only support a subset of possible operations to be done on the encrypted data, i.e. addition or multiplication (but not both). This restriction makes the computations more efficient and allow work with larger datasets in smaller amounts of time.

Somewhat Homomorphic Encryption

schemes, similarly, add limits to the process for the benefits of efficiency, this time limiting the number of operations (additions or multiplications) that can be safely done to a fixed limit. Going beyond that limit leads to data corruption so the limit is fixed based on the specifics of the use case and is agreed beforehand as part of designing the solution.

Fully Homomorphic Encryption (FHE)

is the ideal state that can handle any number of all types of supported operations. It doesn't have the limitations of the other two schemes but is currently prohibitively expensive for processing large data volumes, from both a memory and CPU utilization perspective, when looking to achieve the same strength of encryption as other techniques.

What benefits does it provide?

This mechanism allows complete secrecy of the data to be maintained as the data will not be decrypted while it is outside the control of the data owner. The results of processing are also kept private (even from the data processor) so the risks of unintentional privacy leakage as a result of processing are mitigated.

What are its limitations?

Homomorphic Encryption, when used by itself, is more suitable for use between two parties, rather than between multiple parties as there can be the risk of unintentional data leakage in some multi-party scenarios where the key owner and one data processor could gain access to data from other parties if they colluded. The main limitation with Homomorphic Encryption is the computational intensity and cost of the processing. This limits the amount of data that can practically be used and currently makes it an impractical mechanism for real-time or near real-time processing.

Also, because the data remains encrypted throughout and there may be limits on the types or number of operations that can be performed, the data processor and data owner need to have pre-agreements in place around the structure and content of the data as well as the processing that will take place so that the data processor cannot interrogate or experiment with the data.

What's possible now?

The implementations available are now approaching "product" maturity rather than proof of concept or pilot, but there are still some limitations, especially performance-wise, that mean that certain use cases may be possible but won't be practical. Most of the more mature implementations now support the Fully Homomorphic mechanisms but work best when capped and limited (Somewhat Homomorphic Encryption).

Accenture is using homomorphic encryption to enable companies to derive insights from encrypted real-world, multi-class, industry data *without* decrypting the raw data. Similarly, the technology is being customized by Accenture for companies cooperating on a shared blockchain accounting system or another similar distributed ledger. This has applications where companies have requirements to deal with both privacy and auditability at the same time.

Technique 4

Secure Multi-Party Computation (MPC)

What is it?

Secure Multi-Party Computation (MPC) provides a mechanism that allows a group of parties to share the benefits of combining their data to create useful outputs while keeping their actual source data private from each other. It provides mechanisms for the parties to jointly compute a function or run an operation on their input data without exposing their data. The protocol means that the parties' inputs remain secret, except for what is purposefully revealed by the intended results of the computation.

These technologies support use cases that allow groups of companies to work together to generate outcomes or insights they cannot get alone but where they are not willing or able to share their data directly with each other. It can also help with concerns around input privacy, where each company wants to be sure they are not exposing anything other than what they intend to share.

A simple illustration for MPC is the Millionaire's Problem: two millionaires want to understand who is the richest, but neither want to share their actual net worth with each other nor trust a third party. An MPC scheme could allow a partial calculation of the answer to be done by both millionaires which, when combined, would provide an answer, but on their own would be meaningless.

Characteristics

Controlled Environment?

No, it is specifically designed to run in untrusted environments.

Data Obfuscated?

Yes, it generally protects the inputs and exposes the outputs to each party.

Encryption During Processing?

Yes, only the output data is seen by the parties involved.

Major Variants

The Garbled Circuit

scheme specifically supports communications between two parties, playing sender and receiver roles, and involves representing the computation to be done as a logic circuit (similar to building computer hardware). This circuit is then "garbled" by one of the parties, which encrypts and randomizes aspects of the circuit and then sends this along with their inputs, encrypted in a similar way. The receiver uses a mechanism called Oblivious Transfer to understand how to represent their own data so that it can be combined with the **Garbled Circuit** to create an encrypted output. Both parties then confer to interpret the output without ever having been privy to the inputs of the other.

Secret Sharing

schemes takes a slightly different approach and are intended to be used with groups of more than two parties, where each acts as a peer in partially computing the output. These schemes involve splitting a shared encryption key into many pieces, one per party, in such a way that the pieces when added back together make up the whole key. The pieces are used individually by each party to process their part of the calculation on their data, but each party is unable to interpret any data processed by anyone else either with their partial key or the original shared key. When all the computations have been done, the results can be combined and interpreted by everyone using the original key.

What benefits does it provide?

MPC can be very effective in cases where the trust of the parties, or even their identities, can be difficult to guarantee. MPC specifically deals with scenarios where the parties involved in sharing data may be actively malicious or compromised, and rather than requiring a high level of trust to avoid this like traditional data sharing, MPC is designed to work securely in spite of these situations. Some schemes can provide security even if only one party is behaving legitimately.

What are its limitations?

Computational costs are the main drawback of these techniques, but there are constant improvements happening in this space. MPC also requires a lot of communication between the parties, which can add further latencies during the computation process. Another factor with some schemes is the complexity of representing a business problem as a logical circuit with a compliant structure, which can require some specialist skills. From a security perspective, one point to note is that MPC doesn't protect against "poisoning" attacks, where one of the parties could attempt to maliciously influence the results of queries by another party by intentionally using false or misleading data to intentionally lead to an answer which is not correct (i.e. exaggerating or understating a statistical result to drive another party to draw incorrect conclusions).

What's possible now?

There are a small number of live use cases currently using MPC approaches to solve real-world business problems. There is also a large amount of ongoing research happening in this space. Generally, both two-party and three-party use cases are possible, but the types of computation as well as data volumes should be a consideration.

Accenture is working with semiconductor ecosystem parties to create a trusted, distributed way to share data using MPC and blockchain. Equipment manufacturers need data to deliver better solutions for their equipment, parts and services, and suppliers need to protect their data as well as that of sub-tier suppliers and customer-restricted data (i.e. data related to on-wafer, off-line metrology and integration). While blockchain provides traceability and control of data views, IP issues are so severe that the equipment manufacturer that operates on raw data is reluctant to share data, even if the analytics processing never leaves the network. MPC will be able to solve this problem and enable trust and secure data sharing.

Contact

Teresa Tung
Managing Director–Accenture Labs,
Applied Intelligence Innovation Lead
teresa.tung@accenture.com

David Treat
Managing Director–Accenture,
Global Blockchain Lead
david.b.treat@accenture.com

Jean-Luc Chatelain
Managing Director–Accenture,
CTO of Applied Intelligence
jean-luc.chatelain@accenture.com

Acknowledgments

This paper, and the deep technical and domain expertise presented, was made possible by the contribution of the following colleagues across Accenture Labs, the Dock and Liquid Studios: Malek Ben Salem,Chia Ray Chang, Patrick Connolly, Giuseppe Giordano, Kirby Linvill, Steven O'Kennedy, Luiz Pizzato, Luca Schiatti and Zhijie Wang.

About Accenture

Accenture is a leading global professional services company, providing a broad range of services and solutions in strategy, consulting, digital, technology and operations. Combining unmatched experience and specialized skills across more than 40 industries and all business functions—underpinned by the world's largest delivery network—Accenture works at the intersection of business and technology to help clients improve their performance and create sustainable value for their stakeholders. With approximately 482,000 people serving clients in more than 120 countries, Accenture drives innovation to improve the way the world works and lives. Visit us at **www.accenture.com**

About Accenture Applied Intelligence

Accenture Applied Intelligence helps clients apply new data science and intelligent technology across their business, and into every function, so they can transform their business and achieve new outcomes at speed and scale. Recognized as a leader by industry analysts, the company helps clients create new intelligence using artificial intelligence, machine learning, proprietary algorithms and app-based solutions, all powered by the Accenture Insights Platform. We collaborate with a powerful alliance and delivery network to help clients operationalize within any market and industry with a focus on speed to value. Combining expertise across industries, analytics, technology and design, Accenture is uniquely qualified to drive new business outcomes with precision, at scale. Visit us at **www.accenture.com/appliedintelligence**

This document is produced by consultants at Accenture as general guidance. It is not intended to provide specific advice on your circumstances. If you require advice or further details on any matters referred to, please contact your Accenture representative.

This document makes descriptive reference to trademarks that may be owned by others. The use of such trademarks herein is not an assertion of ownership of such trademarks by Accenture and is not intended to represent or imply the existence of an association between Accenture and the lawful owners of such trademarks.