

Maximize collaboration through secure data sharing



Contents

Introduction	3
What are Privacy Preserving Computation (PPC) Techniques?	5
Where are the opportunities?	9
What's next?	16
Appendix	19

Introduction

Faced with the challenge of competing in an increasingly AI-driven world, enterprises are becoming more aware of the importance of expanding their access to data through third-party partnership ecosystems to create new advantages and opportunities for growth.

The increasingly distributed nature of customer data means many enterprises do not generate the necessary levels of data on their own to derive the unbiased insights required to provide new experiences, open new revenue streams and apply new business models.

This growing need to share data is reflected in a recent Accenture C-suite survey, where 36 percent of executives indicate that the number of organizations they partnered with had doubled or more in the last two years. The same survey also revealed that 71 percent of executives anticipate the volume of data exchanged with ecosystems to increase.

Similarly, according to a recent Harvard Business Review Analytics Services Survey, 78 percent of companies highlighted the ability to easily access and combine data from a variety of external sources as very important for a data-driven enterprise. However, only 23 percent said they were currently very effective in this area, and 15 percent were sharing data with key vendors and suppliers. Why are enterprises failing to unlock the trapped value of their data and that of their ecosystem partners?

While many rightly point to difficulties in overcoming legal and technical hurdles, these challenges are not insurmountable. The major barriers to effective data sharing that exist today include the fact that:

1 Trust remains elusive. In a 2018 EU survey of European companies on B2B data, companies indicated that a lack of control over the usage of data by other organizations (42 percent) was as much of a barrier to data sharing as legal uncertainty over data ownership rights (54 percent). 15 percent of respondents also cited the uncertainty of liability costs in case of damage caused by data sharing as another barrier. Trust between parties is imperative, but traditional trust architectures are incapable of solving this unique problem of data sharing because they do not protect companies against the extended consequences it presents; the risk gap remains too large when considering sensitive data.

2 The risk of sharing data is disproportionately higher than the potential value of sharing data—even in the presence of trust. Accenture research on the impact of revenue on companies after a large, public data breach shows that companies can see an almost ten percent-decline in revenue for up to six months after the breach compared to companies who did not suffer a breach. Worse, it can take almost two years to recover that lost revenue, by which time those companies could be even further behind their competitors.

However, a new family of Privacy Preserving Computation (PPC) techniques, 30 years in the making, are poised to significantly disrupt the enterprise data exchange space.

These techniques are set to address the two key barriers by allowing data to be jointly analyzed without sharing all aspects of that data. By doing so, companies can gain back control of their data and the risks associated with sharing it, even when used beyond their borders.



What are Privacy Preserving Computation (PPC) Techniques?

Privacy Preserving Computation (PPC) techniques are a family of very modern cybersecurity techniques that, instead of focusing on protecting data from access by unauthorized parties, look at how to represent data in a form that can be shared, analyzed and operated on without exposing the raw information. Encryption techniques often form the core of how PPC techniques provide these capabilities, but they are used in a slightly different way than usual.

Traditionally, encryption was used to ensure the security and integrity of sensitive data against unauthorized access while in transit between parties and while at rest. Although encryption provides a reasonable amount of protection from outside interference in transit, in order to process the

data, the data recipient must have access to the keys to decrypt that data. However, there are two risks during this process that should be considered:

1 The risk of compromise: Since the data processor has access to the key to decrypt the data, a breach on their side can compromise the data. Also, since the data is decrypted during computation, there is the chance that the data can be compromised at that point.

2 The risk of trust: Again, because the processor can decrypt and see all of the data, there is a limit on what most businesses are generally willing to share because of a potential loss of competitive advantage. For this reason, the bar for trust tends to remain very high before a company will share data with another at all.

These risks could limit the value businesses can extract from their sensitive data because they hinder potential data sharing collaborations.

PPC techniques use encryption differently to provide a mechanism to share data with other parties while limiting how or where the other parties can access the data, what parts of the data they can see, or what they can infer from the data. There are different schemes adopted by different PPC techniques to achieve this, but they usually do one or more of the following:

- 1 Control the environment within which the data can be operated on**
- 2 Obscure the data to protect the privacy of the data and remove identifying traits**
- 3 Provide a way to allow the data to be operated on while encrypted, i.e. processing it without ever decrypting or seeing the data**

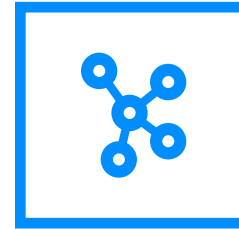
You could think of this as cooking a meal without seeing the ingredients or doing a jigsaw puzzle without seeing the picture of the intended outcome.

Here are some of the primary PPC techniques that are gaining prominence today.



Trusted Execution Environment (Secure Enclave):

An environment with special hardware modules that allow for data processing within hardware-provided, encrypted private memory areas directly on the microprocessor chip only accessible to the running process (page 22).



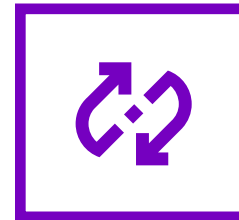
Homomorphic Encryption:

A technology that 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 (page 31).



Differential Privacy:

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 (page 26).



Secure Multi Party Computation (MPC):

A technology that 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 (page 36).



PPC techniques are maturing for analytics and AI use cases

While these PPC techniques and technologies are still new, they are rapidly maturing and are now at a point where they can be used in real business use cases. Securing data at rest, in transit and even during computation is now possible using Trusted Execution Environments. Publicly sharing statistical data without compromising the privacy of individual records is now possible with Differential Privacy. Analyzing encrypted data is now possible thanks to the development of encryption schemes, such as Homomorphic Encryption. Through these technologies, sensitive data can be encrypted and protected at all stages and can be used by a number of trusted or untrusted parties to generate insights without unintentionally exposing data.

With this peace of mind, PPC techniques open many new opportunities for enterprise collaborations that were not previously possible due to risk or regulation.

Beyond the traceability and control of data considerations, these technologies enable partners to work in a decentralized way, giving them the opportunity to jointly investigate common or shared business issues.

Companies are also able to apply AI and improved analysis methods to datasets that they had not previously had access to. This means collaborations with external parties—even competitors—are now possible, and in some cases, well under way.

Where are the opportunities?

Following a long incubation period, PPC techniques are on the cusp of a new phase of industry adoption thanks to the alignment of technological capabilities with market needs.

The potential opportunity has led to the creation of a rapidly-evolving and heavily-funded start-up ecosystem. And innovative enterprises and institutions are investing in and experimenting with these techniques to understand the art of the possible.

What's happening now

For one, Google released its open-source Private Join and Compute protocol this year, leveraging Homomorphic Encryption and MPC. While still at an early stage in terms of enterprise robustness, the protocol serves to illustrate the growing importance of PPC techniques.

PPC techniques are also being used today to help competitors operating in the same market or to allow collaborations in highly-regulated fields, such as drug discovery.

In one of the first commercial implementations of MPC, the Danish Sugar Industry collaborated with Partisia® to develop a confidential production contract exchange amongst sugar beet growers, enabling the industry to readjust to new market situations. Separately, in 2019, ten large pharmaceutical companies developed the Melloddy consortium, which uses blockchain and federated learning to train a drug discovery machine learning algorithm using shared data.

Additionally, PPC techniques are enabling enterprises to develop new, trustworthy data-sharing relationships with consumers. For example, Kara is a privacy-preserving, tokenized data cloud, leveraging trusted execution environments and differential privacy to create a secure way for patients to share and monetize their medical data with researchers, while retaining full control of their data. Kara runs on Oasis Labs' blockchain-based platform and is the basis of a medical trial currently being run at Stanford University. Medical researchers can submit AI systems for training, without ever seeing the underlying data.

PPC techniques are also being used to address many regulatory concerns in markets, such as banking, by accessing and processing sensitive data in encrypted form to derive insights only. For example, ING Belgium uses Inpher's XOR Secret Computing Engine to build analytical models using data from multiple countries like Switzerland and Luxembourg that have stringent data security and personal privacy rules. Proprietary algorithms generated by ING data science teams are compiled with XOR and secretly computed by all regional data centers and/or cloud services providers without revealing any sensitive information; no PII is exported from any jurisdiction.

Furthermore, PPC techniques have already been used by governments. In 2015, the Estonian government worked with Sharemind® to develop the Private Statistics Project, which performed an analysis of a combination of identifiable tax and education records using MPC. The European Commission PRACTICE project analyzed this project and agreed with the Estonian Data Protection agency's findings that no personal data had been processed.

There are even plans to use PPC techniques in upcoming elections: Travis County, Texas is set to implement STAR-Vote—a Secure, Transparent, Auditable and Reliable voting system—which uses homomorphic encryption, to monitor the verified voting process ahead of the 2020 presidential election.

Nascent and emerging opportunities

While the ability to securely share sensitive data presents immediate opportunities, there are also emerging opportunities to disrupt existing markets through the combined effect of PPC techniques and other technologies like blockchain and IoT. Currently, a number of companies and organizations are considering these technologies in areas such as:

1 Consumer data ownership. MyHealthMyData (MHMD), an EU-funded project, is looking at how to share anonymized data for medical care, research and development, all while giving people ownership over their health data. The platform combines blockchain, smart contracts, dynamic consent and a comprehensive suite of data privacy and secure analytics tools including Homomorphic Encryption and MPC.

2 Smart and connected cities. In 2019, it has been reported that Chinese automotive giant, Wanxiang Holding Co., Ltd., and blockchain start-up PlatOn joined together to develop a “smart city” in Hangzhou. As part of the technological development, PlatOn plans to use MPC and other privacy-preserving technologies to help ensure sensitive data, such as residents’ digital identities and information from connected devices, are secure while they interact with one another on a shared ledger.

3 Privacy-preserving data services and marketplaces. The data sharing platform, Ocean Protocol, enables anyone to build data services and marketplaces and provides safe, privacy-preserving and borderless data sharing; it also leverages blockchain and federated learning. Some of the first companies to leverage this network are Aviva and ConnectedLife, who are analyzing—and applying AI to—smart home data to improve care for aging populations; Roche Diagnostics, who is looking to provide better care for patients on blood-thinning therapy; and Next Billion, who is piloting a new data sharing model for small business owners in emerging economies.

4 IoT marketplaces and the Machine-to-Machine Economy. Weeve, a start-up focused on building a decentralized IoT platform, partnered with SingularityNET.io, a full-stack decentralized AI solution provider, to develop AI as-a-service in IoT-based data marketplaces. As part of the partnership, SingularityNET will leverage homomorphic encryption and multi-party computation to support privacy-enabled AI.

5 Decentralized AI. OpenMined is a community that focuses on building open source tools and frameworks that enable the implementation of decentralized AI applications for good. It combines federated learning, MPC, Homomorphic Encryption and blockchain smart contracts to enable decentralized collaboration between enterprises, data owners and data scientists to implement AI applications.

What Accenture is doing

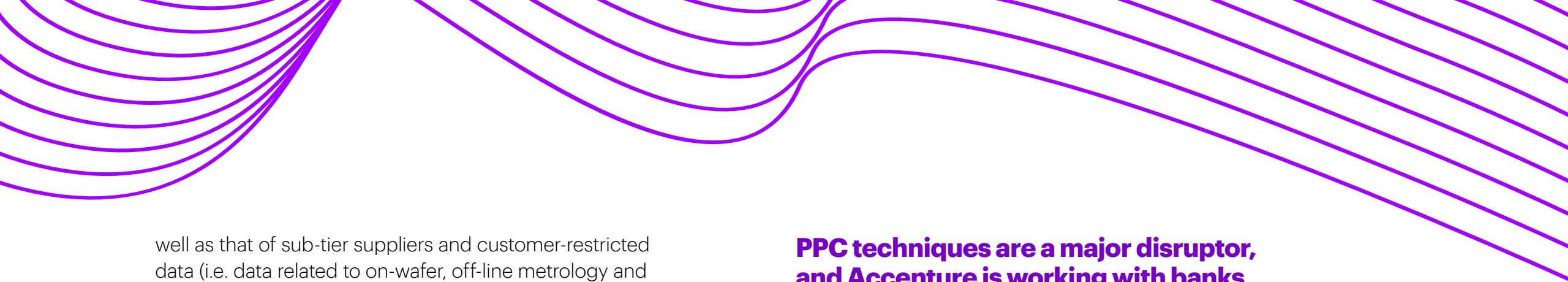
At Accenture, we're working across multiple industries to bring secure data sharing and enable greater collaboration.

For starters, we are calibrating our understanding to assess the various Homomorphic Encryption and MPC frameworks and whether they can address the types of computation each use case requires. In this nascent field, different frameworks specialize in different types of computation—i.e. arithmetic, linear regression, or random forest models—making them well-suited to certain use cases and less suited to others. We are assessing the various trade-offs with each PPC technique, especially in terms of its impact on performance and how new styles of hardware design—like Field Programmable Gate Array chips (FPGAs) and General Purpose CPUs (GPGPUs)—can reduce the impacts on time and cost of using these technologies.

As we identify opportunities, we look at PPC techniques to expand the aperture of available data for AI.

Accenture has been helping companies adopt AI and blockchain—and we're looking at PPC techniques to lower barriers so AI can access more data, including high risk and confidential data.

For example, we are 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 usually reluctant to share data, even if the analytics processing never leaves the network. MPC will be able to solve this problem and enable trusted and secure data analytics.

Specifically building off blockchain, we are customizing PPC techniques for companies cooperating on a shared blockchain accounting system or another similar distributed ledger. This combination has applications where companies have requirements to deal with both privacy and auditability at the same time. We've also recently released a new Open Source project called [PyHeal](#) to help make using some of these frameworks, like [Microsoft SEAL](#), more broadly accessible and adoptable to users in a business context.

PPC techniques are a major disruptor, and Accenture is working with banks to understand how technologies like Homomorphic Encryption could help in cross-border anti-money laundering and anti-fraud use cases.

It could be of huge benefit for banks to be able to ask questions on potentially-fraudulent transactions to other banks without having to expose their customer's data or request data from the other bank, which is often impossible because of banking confidentiality laws and other similar legislation. Technologies such as MPC and Homomorphic Encryption would allow the banks to answer questions on a virtual, shared dataset without the need to share the actual data with each other.

What next?

As discussed, innovative companies have already begun to deploy PPC techniques in real-world scenarios, combining internal capabilities with external PPC expertise. Enterprises looking to embrace these possibilities need to:

Calibrate Understanding:

PPC techniques have a variety of strengths and weaknesses and are currently suited to specific applications and data requirements. Companies should develop a calibrated understanding of what is realistic today, where these technologies can be applied in their industry and which PPC techniques are fit for their specific goals.

Identify Opportunities:

Companies should work with ecosystem partners to identify previously inaccessible data-sharing opportunities, or look at existing high-risk, cumbersome data-sharing processes that may be suitable for initial test and validation. In particular, AI drives these opportunities with companies looking to gain clearer insight via wider access to data beyond what each can provide on its own. Because PPC techniques operate without the raw data ever being revealed, they lower the barrier to data use.

Co-Create:

Companies should look to foster innovation and co-create with trusted partners that have expertise in the area of PPC techniques and tangential technologies like blockchain and cybersecurity to help improve existing tools and technologies and to fashion new tools in these areas that we don't yet know we'll need.

Think long term:

With further development, PPC techniques have the potential to be a major disrupter of traditional data-driven business models while also democratizing the raw materials of AI as they can support decentralized algorithm development rather than being dependent on a few large AI vendors. Companies should start to think long term, investigating opportunities to use these techniques in conjunction with other technologies to create new value.

With the growing argument for ecosystem collaboration and the need for effective and secure methods of data exchange and collaboration, PPC techniques and their successors will be critical to effective, safe and secure data sharing.

This in turn will be fundamental for businesses to gain value from their data and provide new avenues for disruption. We're on the path to industrializing PPC offerings so our clients can take advantage of new near- and long-term opportunities.

Let's talk about what that could mean for your business.

Appendix

Glossary of terms	20
Trusted Execution Environment (Secure Enclave)	22
Differential Privacy	26
Homomorphic Encryption	31
Secure Multi-Party Computation (MPC)	36

Glossary of terms:

Data Obfuscation:

The process of hiding the original data by modifying the content, i.e. by replacing certain parts of the content with meaningless content while keeping the data usable. Usually used to protect sensitive or personally-identifiable data and is also referred to as Data Masking.

Anonymization/De-identification:

Refers to types of obfuscation that are intended to maintain privacy by replacing personally-identifiable content, i.e. names, addresses, phone numbers etc., with values that don't have direct relationships to that person.

Internet of Things (IoT):

Refers to the embedding of systems and sensors into physical devices and objects that allow them to interconnect with each other and to the wider internet without human intervention.

Field Programmable Gate Arrays:

Types of microchips that allow their own internal configuration to be configured by end users or integrators that allow the chips to be set up in a way that is tuned and tailored to the exact use case of the owner. This allows greater performance to be achieved by using hardware tailored for a given purpose without the need to commission custom hardware.

General Purpose Graphics Processing Units (GPGPUs):

Hardware chips specifically designed to handle the highly parallel tasks of rendering and refreshing complex (2D and 3D) graphics on a screen. It was found that these same chips were much more efficient than standard CPUs at doing other processing tasks with parallel loads. GPGPUs are an extension of the same types of chip but tailored further towards more general parallel data processing than graphics processing.

Federated Learning:

An approach to machine learning for training a central, shared model using data that is distributed across multiple locations rather than available centrally. It has applicability where all the training data is not available in the same place or at the same time or where it is not possible or desirable to bring the training data into a central location. It allows the data to be used where it exists without the need to remove it from its location (i.e. a mobile phone or other device) and then uplinks the learnings back into the central model without having to send the actual data back.

A more detailed look into PPC techniques

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?

Trusted Execution Environments (TEE) offer hardware-based isolation for specific application code and data and are designed to provide protection from processes running at higher privilege levels. TEEs can be used to process data between untrusted parties.

What are its limitations?

A TEE can provide confidence in the confidentiality of computation and outputs but not in the integrity of initial inputs. TEEs are only available on certain hardware sets.

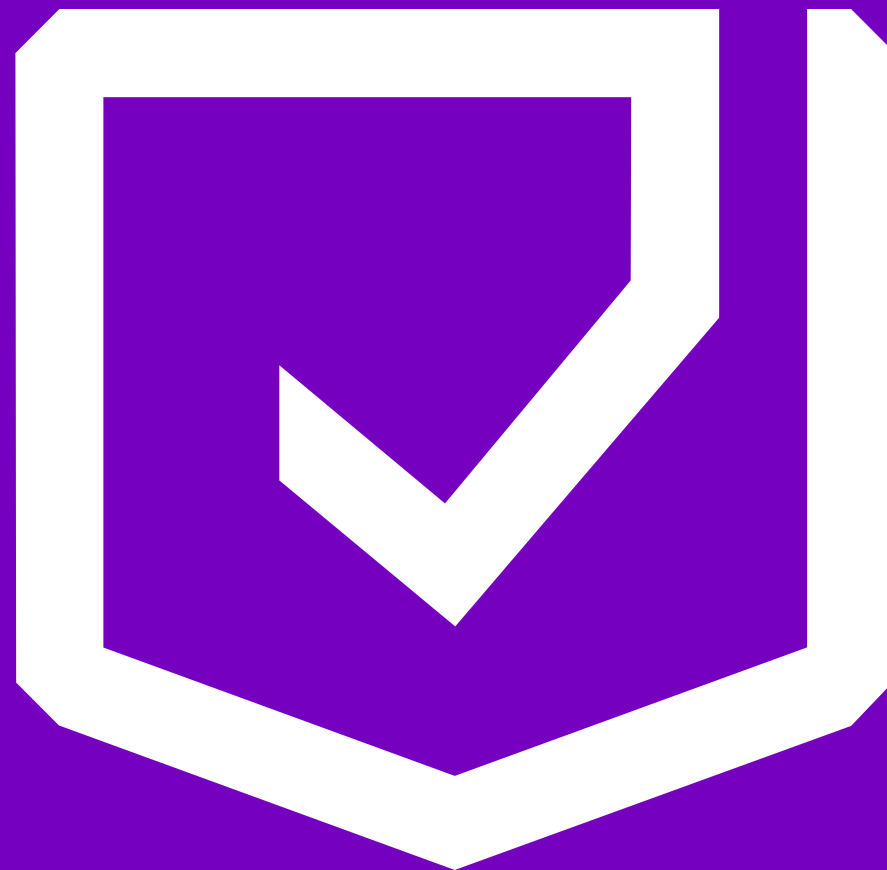
What's possible now?

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 SnowflakeSM 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?

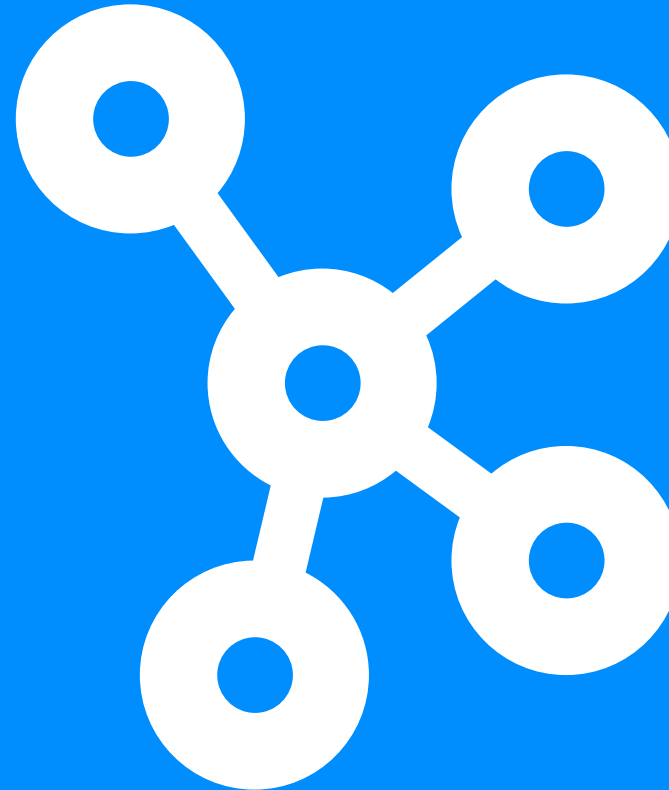
It is only useful for statistical processing or data aggregation as the noise added will change the real data but keep its statistical properties. Therefore, it will not work for use cases where the data integrity of specific values within the records is a fundamental part of what needs to be shared.

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.

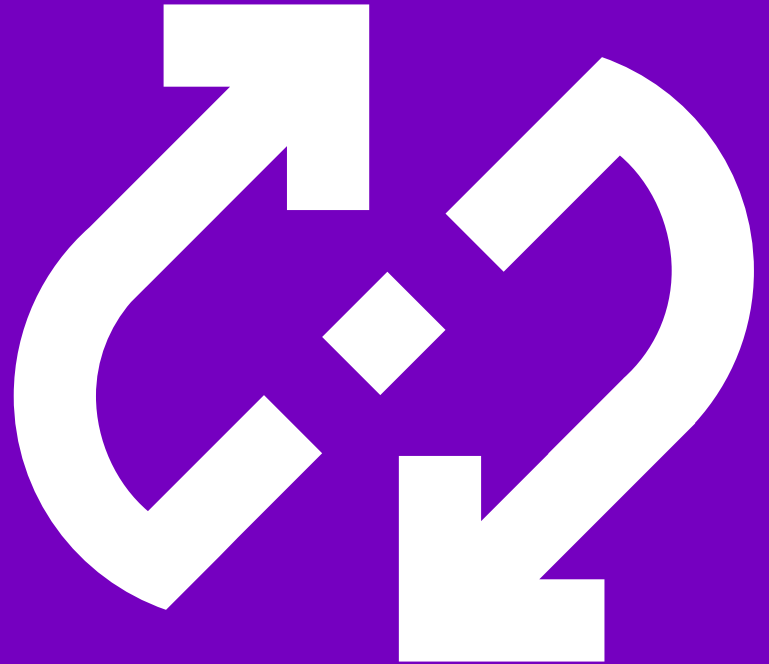
What's possible now?

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

Patrick Connolly

Research Manager–The Dock,
Accenture Research
patrick.connolly@accenture.com

Copyright © 2019 Accenture. All rights reserved.

Accenture and its logo are trademarks of Accenture.

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 492,000 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

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, Giuseppe Giordano, Kirby Linvill, Steven O’Kennedy, Luiz Pizzato, Luca Schiatti and Zhijie Wang.

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.