Federal Register Notice 87 FR 35250, https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-for-information-on-advancing-privacy-enhancing-technologies, June 9, 2022

Request for Information (RFI) on Advancing Privacy Enhancing Technologies

Intel Corporation

DISCLAIMER: Please note that the RFI public responses received and posted do not represent the views or opinions of the U.S. Government. We bear no responsibility for the accuracy, legality, or content of the responses and external links included in this document.

Intel's response to RFI 35250 Federal Register / Vol. 87, No. 111 / Thursday, June 9, 2022 / https://www.govinfo.gov/content/pkg/FR-2022-06-09/pdf/2022-12432.pdf

July 8, 2022

Intel Corporation (Intel) appreciates the opportunity to respond to the Request for Information for Advancing Privacy Enhancing Technologies (PETs) published by the Federal Register at https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-for-information-on-advancing-privacy-enhancing-technologies. Intel agrees with the ideas on Privacy Enhancing Technologies expressed by OSTP in the recent blog.

Intel's mission is to engineer solutions to address society's greatest challenges, providing our customers with reliable, cloud-to-edge computing, inspired by Moore's Law. Intel designs and manufactures advanced integrated digital technology platforms that power the connected world.

In this response we provide an introduction on PETs and address the following points in the RFI:

- Specific technical aspects or limitations of PETs.
- Specific research opportunities to advance PETs.
- Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs.
- Risks related to PET adoption.
- Existing best practices that are helpful for PET adoption.
- Other information that is relevant to the adoption of PETs.

Intel has researched and created PET's technologies for decades, including, but not limited to, research and standardization in fully homomorphic encryption, work in Multi-Party Security Computation and Federated Learning, and development of PETs friendly standards protocols, e.g., Direct Anonymous Attestation (DAA).

Introduction

According to a common definition, "**[p]rivacy-enhancing technologies** (**PET**) are technologies that embody fundamental data protection principles by minimizing personal data use, maximizing data security, and empowering individuals and institutions to make decisions based on their data[1]." PETs identify systems, processes, and techniques that enable processing to derive value from data while minimizing privacy and security risks to individuals and institutions.

PETs deal with protecting the sensitive part of the data in an information system, e.g., private information about persons, institutions, or governments, through techniques such as data obfuscation (e.g., information theoretical metrics privacy, use of synthetic data), data confidentiality (with cryptography and security mechanisms), system architectures for federated

[1] https://en.wikipedia.org/wiki/Privacy-enhancing_technologies

analysis (e.g., secure multi-party computation, federated learning), and a combination thereof, without sacrificing the usability of the data.

PETs-based solutions help individuals and institutions to develop tools that assist with compliance with data protection regulations, enhance data integrity, mitigate cyber security risks, and preserve confidentiality while enabling data analysis and other functionality. PETs can be used in tandem with governance and operation systems capabilities to realize their benefits.

It is also important to note that PETs comprise a large portfolio of tools and techniques that are at different degrees of maturity, with some approaches already widely deployed, while others are in early stages of research. As we describe these technologies, the level of maturity will be specified if relevant. Greater investment in research in PETs can benefit both mature and emerging technologies in this area.

Specific technical aspects or limitations of PETs

PETs address privacy and data protection in a variety of ways. Important state-of-the-art techniques in this area include foundational technologies such as differential privacy, use of synthetic data, fully homomorphic encryption and other cryptographic mechanisms, secure multi-party computation, and federated machine learning, as well as hardware foundations for PETs, such as Confidential Computing and Zero-Trust technologies. We briefly explain some of these techniques below:

- Differential Privacy (DP): DP adds random data, or "noise," to the real data, such that real information is hidden in the noise. DP allows for accurate analysis to be done on data in aggregate. The amount of "noise" is associated with the ability (or difficulty for an attacker) to re-identify the data. Hence, institutions using DP can set a privacy "budget" to tradeoff the difficulty of re-identifying data with the utility of the data.
  Experiments in several use cases show that adding reasonable privacy guarantees does not meet utility requirements. Additionally, differential privacy guarantees depend heavily on the privacy objectives of the solution. Further work is required on DP mechanisms to lower the impact on performance, especially when being utilized to protect large scale AI models. Research in this area focuses on alternative empirical metrics, a shift to localized data analysis, compression methods to minimize data movement, as well as information theoretic methods to protect certain attributes of data.
- Synthetic Data (SD): SD replaces real data with altered or artificially generated data. It inserts additional, synthetic (not real or measured) information into real datasets. Synthetic data are commonly created through machine learning and mimics the characteristics of real-world data.
  The data are generated by feeding real data into machine learning algorithms, which then identify data characteristics and trends, and replicate them in the form of synthetic data. Using synthetic data limits the need to collect and store real information about individuals. A main drawback is that the process of creating synthetic data using AI is not yet mature.

Additionally, the study of the limitations of synthetic data in comparison to real data is emerging, and a more thorough examination is needed to ensure adequate accuracy in analysis based on synthetic data.

- Federated Learning (FL): FL allows multiple owners of private data sets to jointly train a model based on the union of their data without sharing their data with other parties. First, a central server shares an initial version of the model with the data owners. Next, each data owner uses its private data to adjust and improve the model. Such adjustments are aggregated at the central server, which then sends the improved model to the data owners for further rounds of adjustments.
  The technology has been applied successfully to a variety of problems (e.g., in healthcare[2]) but continues to be a subject of active research to expand its scope to allow data sharing and learning at a massive scale, with new mechanisms being devised to make it operate with inherent resource limitations in emerging applications such as networked IoT/immersive media applications.
- Secure Multi-Party Computation (SMPC): SMPC enables two or more parties to compute an output that depends on inputs that each party would like to keep secret, in a way that the parties learn the output but nothing about the secret inputs.
  SMPC has been successfully deployed for key management in IT infrastructures, in point solutions in the finance, cryptocurrency, and code signing domains. There are many MPC cryptographic protocols, but they all tend to be communication intensive, that is, rate-limited by communication bandwidth between the parties, and memory footprint increase with the complexity of the computation.
- Cryptographic primitives and Fully Homomorphic Encryption (FHE): Encryption techniques are essential tools to protect data confidentiality. Classic encryption techniques (such as symmetric and asymmetric cryptography) protect confidentiality when data is temporarily not in use (e.g., at rest, and in transit). The shift to Post-Quantum Cryptography (PQC) driven by the advent of Quantum Computers and the NIST selection of candidate algorithms to standardize, urges to increase the key size used in symmetric cryptography (to resist to the Grover's algorithm), and to deploy quantum-resistant public key cryptography, e.g., lattice-based cryptography, for key encapsulation and key agreement (to resist the Shor's algorithm). Introducing quantum-resistance in future cryptographic suites implies an increase in cryptographic overhead, which in turn may favor hardware assisted acceleration to pure software solutions.
  Modern cryptographic techniques, such as Fully Homomorphic Encryption (FHE) make it possible for data to be operated on and modified while in encrypted form and without having to disclose the associated decryption keys. However, among the implementations of FHE schemes to date, none are considered computationally efficient on existing hardware architectures to be widely utilized in commercial products.

[2] https://newsroom.intel.com/news/intel-works-university-pennsylvania-using-privacy-preserving-ai-identify-brain-tumors/

Unlike classic cryptography, FHE is still in its early stages of maturity from a deployment perspective, but it has the potential to be used widely in applications ranging from smart contracts to payment processing as the technology advances to become progressively less expensive.

FHE has several important properties unlike other methods used to process encrypted data. First, the deployment of FHE within a data processing pipeline does not disrupt the data lifecycle. As a result, when FHE is deployed in AI/ML data lifecycles, it can either preserve or amplify the value of data while reducing the risk of third-party data leakage during processing. Second, FHE can be deployed in various forms, including shared key, public keys, or secret sharing; this allows flexible deployments for different usages to elevate the level of confidentiality in existing security solutions while reducing the risk of third-party data leakage. Third, when FHE is deployed, the physical location of the servers where data is processed can be anywhere on earth or in space. Fourth, FHE can be deployed at both classical and post-quantum security NIST levels. Fifth, data can be encrypted before, during and after processing, providing protection throughout its lifecycle. In summary, using FHE can elevate the level of confidentiality of the security solution by preserving both data privacy and the value of data. At the same time, FHE can enable considerable business and public service opportunities through data sharing and collaboration.

The main challenges with homomorphic encryption include: the resources required for deployment, the lack of widely accepted standards, and practices to integrate FHE with other PETs and IT infrastructure components. Despite this, there are practical yet specialized uses of homomorphic encryption today. For example, FHE is being used to monitor whether passwords saved in Web browsers (e.g., Microsoft Edge) were ever exposed in a data breach.[3]

- Confidential Computing as hardware foundation for PETs: Confidential Computing is the protection of data in use by performing computation in a hardware-based Trusted Execution Environment (hereby TEE). A TEE is a key technology for data protection is a Trusted Execution Environment. A TEE is an environment that protects code and data confidentiality and integrity during execution. A hardware-based TEE typically provides these properties by mapping the TEE to a secure portion of memory and enforcing access-control from non-TEE executions. In addition, the hardware provides abstractions for attesting the state of a TEE to a third party that allows an external verifier to ensure that execution is happening in a TEE to gain trust. Developers can host portions of their applications into hardware-isolated programs or hardware-protected areas of execution to increase security, even on compromised platforms. Using TEEs, developers can protect the integrity of sensitive software and crucial configurations. Many cloud providers already offer hardware assisted trusted execution as part of their services, and companies can further increase protection by looking into third-party encryption solutions and providers. The technology is mature, so many of these hardware features are available today for both

[3] https://www.microsoft.com/en-us/research/blog/password-monitor-safeguarding-passwords-in-microsoft-edge/

businesses and individuals, and accessible mainly thanks to the open-source efforts to deploy unmodified code in a TEE.

It is important to highlight that TEEs can enforce not only confidentiality, but also integrity of workload memory (typically using a combination of both access-control and cryptography mechanisms) even in the presence of privileged malware at the system software layers. TEEs in different forms have been the subject of standardization for two decades and are mature and adopted solutions.

- **Zero Trust and Trust as a Service (TaaS) as foundational service for PETs:** TEEs' functionality can be enhanced with the use of Zero Trust architecture. This type of architecture is based on the principle that all the resources must establish their trustworthiness before they are utilized, from the provenance of the platform to the execution of sensitive workloads in the infrastructure where the workload owner may have little control. Zero Trust architecture can support and supplement capabilities of PETs. Attestation of trustworthiness supports principles of Zero Trust and builds a stronger security foundation for PET-based systems.

  Intel's vision of Trust as a Service establishes independent trust authority to attest to every computing asset in public cloud, edge, and private cloud. Trust as a Service (TaaS) can provide proof of trustworthiness within the key areas described below, strengthening the foundations of PETs or incorporating some features of PETs:

  1. The platform has established correct provenance from the earliest point in the supply chain or software/data life cycle.
  2. The platform has foundational integrity (for example, in its firmware).
  3. The software on the platform has not been tampered with.
  4. The integrity of the TEE (Trusted Execution Service) has been verified.
  5. The relevant policies are enforced.

Specific research opportunities to advance PETs

With all the pros and cons about PETs, it is clear that the following types of initiatives have had the largest impact on advancing PETs:

- The establishment of public-private partnerships, e.g., funded by the US government through DARPA and IARPA (e.g., the DARPA PROCEED (archived), IARPA HECTOR (archived) and the DARPA DPRIVE[4] (active) programs to advance FHE) and the National Science Foundation (NSF) (e.g., the partnership on "Resilient & Intelligent Networking Systems" (RINGS)[5], and "Machine Learning for Wireless Networking Systems" (MLWiNS)[6] to advance AI practices).
- The establishment of industry driven academic partnerships, e.g., led to the creation of the Gramine platform[7] for wider adoption of hardware-based trusted execution by deploying unmodified applications.

[4] https://www.darpa.mil/news-events/2021-03-08
[5] https://www.nsf.gov/pubs/2021/nsf21581/nsf21581.htm
[6] https://www.nsf.gov/pubs/2019/nsf19591/nsf19591.htm
[7] https://gramineproject.io/

- The establishment of industry consortia and corresponding open-source initiatives to advance practices and uses of PETs and their hardware foundations, including but not limited to the formation of the Confidential Computing Consortium[8] for hardware-based TEEs; the formation of the MPC alliance[9] for SMPC; and the Open Federated Learning project for wider adoption of federated machine learning[10].
- The development of standards, e.g., around FHE[11] and SMPC[12]; the NIST selection of post-quantum cryptography (PQC) candidates to standardize advances in the field by allowing institutions to accelerate the adoption of PQC algorithms in products.

Specific sectors, applications, or types of analysis that would particularly benefit from the adoption of PETs

PETs can be and are applied in a variety of fields and use cases, e.g., AI modelling, cross-border data transfers, and data analytics to help security and risk stakeholders manage constraints while respecting individual privacy, and ultimately reduce direct (by direct access) and indirect (as a consequence of processing) third-party data leakage. Examples of applications include, but are not limited to:

- Healthcare: information including patient data can be pooled together for research, drug-development, and public health. The COVID-19 pandemic has highlighted the value of information in public health, from contact tracing to vaccine development. Given the importance and particularly sensitive nature of data in healthcare, this sector has been a leader in exploring the potential of PETs in emerging technology areas, such as FL[13] and FHE[14].
- Financial services: requirements such as Know-Your Customer (KYC), credit reporting, money laundering detection, and fraud mitigation drive the collection of data that is matched to specific individuals. Pooling this data together can help identify financial crimes and protect individuals from financial loss.
- Government applications: processing of super-resolution images; localization.

Since PETs cover data privacy and confidentiality, data aggregation use cases for PETs are not limited to situations where privacy of data subjects needs to be protected. They can also be used in data confidentiality cases, including information sharing in cybersecurity and other areas, such as intellectual property protection (e.g., AI models).

[8] https://confidentialcomputing.io/
[9] https://www.mpcalliance.org/
[10] https://github.com/intel/openfl
[11] https://www.iso.org/standard/83139.html
[12] https://www.iso.org/standard/80508.html
[13] https://newsroom.intel.com/news/intel-works-university-pennsylvania-using-privacy-preserving-ai-identify-brain-tumors/
[14] https://carnegieendowment.org/2021/08/17/introduction-how-korea-can-unleash-power-of-data-pub-85162

Risks related to PET adoption

As in all technology areas, there are several challenges associated with the adoption of PETs that need to be overcome for successful adoption:

- **Internal capacity and expertise within organizations to deploy and manage PETs**. This is a common issue across technical deployments that require specialized expertise, but **since many current generation PETs are not widely used, their broad deployment can pose unique challenges.** Some techniques or systems may be able to function with limited support, while others need more oversight. Some technologies may also need to be used in conjunction with business enterprise systems, and therefore, require work to integrate and maintain those connections, which can increase costs.
- **PETs are in various stages of maturity**. While promising, some techniques and systems are still in early phases of development and require additional research before they are commercialized. The variability in maturity and research adds to the complexity of PET adoption and may make it harder for institutions to determine which PETs are appropriate and what resources they need to deploy them.
- In the systems that we have examined **it can be difficult to validate correct implementation of PETs.** As in the case of cryptography, some current PETs depend upon code review to ensure correctness, as correctness of privacy preservation cannot be validated by examining the outputs. This motivates the need for validation techniques and tools, such as heuristic measurements, to compliment the PETs themselves. It also motivates the value of standards and standard implementations such as cryptography standards.

Existing best practices helpful for PET adoption

The new techniques described above are broadly applicable to diverse datasets with multiple owners. While there are some existing standards which can support privacy while preserving data processing in these contexts, new standards, best practices, and guidelines will need to be created. This is an important area of pre-standardization research and development. Hence, best practices, including but not limited to the following, remain helpful for PET adoption:

- Transitioning to practice the results of research and development outcomes of public-private partnerships.
- Adopting the outcomes of best practices and usages developed by reputable industry consortia.
- Transferring the technical results of industry driven academic partnerships into industry, standards, and best practices.
- Supporting open international standards in the area.
- Building PETs on a secure foundation using up to date security models, e.g., Zero-Trust, secure hardware.
- Ensuring high-quality data and adequate data governance.

Mechanisms, not covered above, that could be used, modified, or introduced to advance PETs

All of the PETs described above aim to address data confidentiality in various forms, and at different stages in the data lifecycle. However, integrity of the data, the computation, and the context is an important property that the integration of PET technologies in actual systems must support.

Computational integrity and attestation are important mechanisms for enforcing privacy policies on data. For many of the PETs listed above it is important to be able to validate that the technology is being correctly applied, since it is not currently possible to determine from the outputs whether the PET was instantiated. Systems that deploy PETs with integrity and attestation capabilities enable remote parties to have confidence that those PETs were correctly deployed and invoked.

Contextual integrity is a privacy concept in which the individual has the right not to be observed or disturbed when they don't want to be. Enforcement of contextual integrity on data use is another dimension of privacy that technology should enable – many privacy violations are contextual integrity violations of data that was accessed without violating confidentiality technologies. Through Confidential Computing, attestation, smart contracts, and, in general, verifiable computation, the beginnings of technical mechanisms to enforce contextual integrity policies on data can be constructed.

To conclude, Intel appreciates the opportunity to respond to the Request for Information for Advancing Privacy Enhancing Technologies. Please do not hesitate to contact Ro Cammarota, Intel Principal Engineer and Claire Vishik, Intel Fellow for further information.