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

Duality Technologies

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.

Duality Technologies Response to OSTP Notice of Notice of Request for Information on Advancing Privacy-Enhancing Technologies

Dr. Ahmad Al Badawi, Senior Scientist
Ronen Cohen, VP Strategy
Dr. David Bruce Cousins, Director Duality Labs
Dr. Nicholas Genise, Scientist
Dr. Kurt Rohloff, CTO and Cofounder

Organization Type: Industry

July 6, 2022

Executive Summary

Duality Technologies Inc., a US-based corporation, thanks the OSTP for the opportunity to contribute the following comments regarding the request for information on advancing Privacy-Enhancing Technologies (PETs).

Duality is a leading PETs provider, enabling organizations to collaborate on personal data or other sensitive data without anyone but the providing data controller having unencrypted access to the data. Our software platform leverages several PETs, including Homomorphic Encryption (HE), Secure Multiparty Computation (SMPC), Federated Learning (FL), and more, and utilizes PALISADE, an open-source, standards-compliant Homomorphic Encryption library built with US Government funding. Duality's founding team is comprised from world renowned cryptographers, including Turing Award winner Prof. Shafi Goldwasser, and data science experts. We enable sharing of data with one or more entities processing the data in a robustly encrypted, fully private manner that protects the data from access by the processors or any other party except for the data controller who encrypts the data.

1. Specific research opportunities to advance PETs

There are several research opportunities that can be addressed to advance the current state of PETs. Following is a non-comprehensive list of opportunities, however we think the below are both immediate and crucial to the advancement of PETs.

- **Utility, privacy, and performance tradeoffs**
  There is almost always a tradeoff between data utility, data privacy, and performance that should be considered in defining the best solution for a use case in question. Maximum utility requires full access to the original data without any added noise, masks or concealments (such as encryption), but this can violate data privacy requirements. Thus, several PETs apply some transformation to the

original data before sharing it for computation. This transformation can affect the extent to which data can be utilized. For instance, differential privacy-based solutions transform data by adding noise that renders the data subjects hard to re-identify, yet the transformed data remains useful for performing some computations like aggregate queries. With this technology, the added noise can affect the precision of some queries, which creates complications for high-precision applications. On the other hand, there exist PETs that apply noise-free transformation on the data (such as homomorphic encryption or multi-party computation), while allowing arbitrary computation to be performed at almost zero loss of precision. While initial realizations of these PETs were impractical, they are currently considered practical, and in some particular cases, very efficient. Further research is still needed in this domain to make them more efficient for arbitrary computations. As an example, this is an active research area for DARPA, who have funded several PET based research programs in the last ten years, including the Cooperative Secure Learning (CSL) program for PET-based machine learning, and the Data Protection in virtual Environments (DPRIVE) program to accelerate the performance of fully homomorphic encryption (FHE) with novel hardware, both of which have Duality as a participant [1] [2].

- **Quantifying privacy for risk assessment**

  One of the main issues in most PETs is their lack of rigorous security proofs. In general, such proofs stem from the computational hardness of well-defined and well-studied mathematical problems like integer factorization, discrete logarithm, and some lattice problems. Cryptographic systems based on these techniques increase the effort an attacker needs to compromise the security of the system to a point where unrealistically large computers would need to run for years or even decades to crack them. In non-cryptographic PET systems, this effort cannot be quantified in the same way. Despite this complexity, quantifiable analysis is necessary for many of today's data-centric applications, as well as to satisfy the requirements of many information and data security organizations.
  The literature is rich with several studies and incidents that showed the weakness of approaches with non-quantifiable security, and how they should be avoided [3] [4] [5] [6]. Yet, these PETs are still being used nowadays in several applications and their usage is even recommended in some data protection acts [7]. In contrast, more robust and concrete PETs such as FHE and SMPC have rigorous mathematical proofs of quantifiable security. The security level of these technologies is well defined and is usually expressed in the maximum number of elementary operations an attacker must do to certainly compromise the security of the underlying system. Therefore, an area of paramount importance to the advancement of PETs is to define and quantify different degrees of privacy in terms of the effort needed to compromise them, standardize these definitions, and focus research on applying them to non-cryptographic PETS, an effort that will make quantitative risk assessments more feasible, reliable, and more relatable to non-technical users.

- **Privacy in light of emerging technologies**
  Research should be furthered to advance existing PETs while addressing arising challenges brought on by emerging data-driven technologies such as 5G/6G, Artificial Intelligence, Big Data, IoT, and Blockchains. Take Artificial Intelligence, for example. Over the past decade, a number of scientific branches in AI, such as machine learning and deep learning, have made huge leaps in numerous data analytics tasks. Fundamentally, these systems are trained on huge amounts of data used to develop models that can make informative predictions or take actions when deployed. Although existing data protection and PETs can be used to enforce privacy in the training phase, it has been found that this can be more challenging when it comes to the deployment phase. While still secure enough to prevent unauthorized access to the underlying data or model, attacks such as membership inference allow passive adversaries to learn whether a deployed model has been trained on a designated data subject or not [8]. This is an example of the kind of attack that can be difficult to predict while developing new PETs that can compromise both privacy and security.

- **The right PETs for the right job**
  We can comfortably say that there is no one-size-fits-all PET that can solve all privacy problems, and as such, this requires a thorough analysis of any targeted use case in order to define the best viable option. Moreover, a PET that might work in one application may not be suitable for other applications. In fact, some use cases might require the employment of multiple PETs within one solution. This area has recently been an area of attention in the PET space; for instance, SMPC and HE have been used to devise more efficient multi-party HE schemes [9] and differential privacy has been married with HE schemes to alleviate some limitations of HE-based inference applications [10].

2. Specific technical aspects or limitations of PETs

As stated in the previous section, different PETs are often best suited for different applications, and may in fact work together in a complementary manner. The two PETs with broadest applicability in terms of functionality are FHE and multiparty computation, since they can securely compute mathematical functions on encrypted data. Therefore, one can build systems to implement any computer algorithm securely using FHE, SMPC, or a combination thereof. FHE allows all portions of a computation to be done on data at rest through computationally intensive cryptographic transformations, whereas SMPC requires passing multiple messages between all parties throughout the computation. Historically, both of these approaches had inefficiencies (compute-limited or bandwidth-limited) that impacted the feasibility of using these PETs for certain applications, although recent advances have improved performance by orders of magnitude. Ultimately, FHE combined with hardware acceleration could be an effective and provably secure solution for most applications despite the initial expense of specialized hardware. Additionally, the combination of FHE and SMPC together is a new research area where the best of both approaches can be combined for better performance. Both areas are worthy of further research, since their security is the strongest among PETs.

Differential privacy (DP) is a statistical encoding that requires the input to conform to a known distribution *a priori*. This is a strong restriction on the input, but makes it efficient since the encoding is usually simple. Unfortunately, if the full statistics of the data are not known in advance, DP could make it difficult to uncover previously unknown relationships in the data.

A trusted execution environment (TEE) is a physically isolated execution environment, usually within a CPU. TEEs are very efficient, and multiple generations of the technology have been developed by various CPU vendors. However, active research has been able to uncover numerous security flaws due to side-channel attacks, so caution needs to be used when implementing systems with this technology in order to avoid these issues. [11]

Other forms of PETs, such as federated learning (FL), synthetic data, and anonymization have significant security drawbacks and are unlikely to pass a rigorous standardization process (e.g., NIST) for broad functionalities. They can, however, carefully be used in limited settings, or alongside other PETs. For example, FL suffers from a reverse-engineering flaw since the model is known to all collaborating parties, and synthetic data or anonymization eliminates the possibility for correlation studies among different datasets. This is why the combination of FL and FHE has been a research area for DARPA and others (for example, as part of DARPA’s Cooperative Secure Learning (CSL) program, which was referenced above).

As mentioned previously, a limitation across all PETs is the lack of a shared set of threat models and a shared, quantifiable measure of security. This can be overcome with more research and a standardization thrust. Lastly, the best long-term solution for most applicable problems may be achieved by mixing PETs under these unified models.

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

There are numerous challenges and use cases in both the public and private sectors that would benefit from the adoption of PETs. PETs are designed to enable the use of data and derivation of insights in challenging circumstances, such as when competitive or proprietary data concerns exist; when regulated data must be used (e.g., HIPAA-protected data, or financial data); when there are privacy laws and frameworks that must be adhered to (e.g., CCPA); when cross-border data transfers are needed in a way that respects the law of each country (e.g., when data localization / data residency laws exist), and many more. Some examples are below:

- Public Sector

    - Investigating crimes while protecting investigation integrity (e.g., for money laundering, counter terror financing, tax fraud, corruption, and more)
    - Secure distributed control of Critical Infrastructure, Networks, IOT
    - Processing and/or distributing sensitive information in a military or security context
    - Collaborating with the private sector in a Public/Private Partnership setting
    - Cross-border collaboration with international counterparts and allies
    - Collaborating and processing data in Zero Trust environments
    - Analyzing sensitive data for public health scenarios
    - Understanding provenance and quality of manufactured items to enable better maintenance / fleet management, and protecting supply chain and logistics
- Private Sector - Healthcare, Pharmaceuticals, and Life Science
    - Accessing, linking, enriching, and analyzing multi-center data to better understand and develop treatments and drugs
    - Leveraging Real World Data for clinical trials and drug development
    - Clinical-Genomic Analysis - Joining genetic information with clinical data to find correlations
    - Health and lifestyle insights leveraging data from wearables
    - Conducting patient recruitment for trials
- Private Sector - Financial Services
    - Cross-institution collaboration to prevent, detect, and investigate financial crimes (including money laundering, terrorist financing, and predicate offenses like fraud, cyber crimes, human trafficking, etc.)
    - Cross-border collaboration (internal to a single institution) on financial crimes, customer service, marketing, product development, etc.
    - Third party collaboration for personalized marketing and service development (e.g., for granting credit)
    - Offering personalized rates and services based on consumer activities (e.g., insurer offering better rates based on driving habits)
    - Benchmarking across organizations (e.g., around cyber security posture)
- Private Sector - Telecom
    - Securing 5G and other critical infrastructure software to run on untrusted hardware (note, DARPA is funding research in this area) [12]
    - Acting as a secure access data broker (for publish/subscribe scenarios)
- Private Sector - Other
    - Collaborating across retailers to better manage inventory and supply chain
    - Collaborating across manufacturers and vendors to increase quality
    - Enabling secure computations in the cloud

4. Specific regulations or authorities that could be used, modified, or introduced to advance PETs

This response combines both the question on regulations (#4) as well as on specific laws (#5). There are several aspects to consider regarding these topics.

First, regulators and lawmakers should define priority areas for collaboration and give organizations a “safe harbor” for collaboration. Good candidates for priority areas may be “public good” use cases like stopping fraud and cyber crimes, money laundering, terrorism, and even enhancing pharmaceutical or medical research.

Second, regulators and lawmakers can be more forthcoming on how and when PETs can enable collaboration in compliance with privacy and industry-specific laws. This also includes a few different facets:

- Harmonizing approaches and interpretations across regulators. As an example, the US alone has over a dozen different financial services regulators at the federal level, and each state often has its own authorities as well. A single financial institution may be regulated by several of them, and all may interpret the law differently, or issue guidance which is unharmonized. This makes it challenging to remain compliant. As an example, in December 2020, FinCEN issued a clarifying remark on information sharing under the PATRIOT Act section 314(b), which they interpreted as including “predicate offenses” to money laundering and terrorist financing (like Fraud, for example) [13]. Unfortunately, other regulators have not formally issued this same guidance, and as such, fraud information sharing under 314(b) amongst US financial institutions remains under-utilized.
- Harmonizing approaches and interpretations across borders (e.g., with US Allies). Many industry stakeholders have multinational operations, and a lack of harmonization across the jurisdictions in which they operate adds to complexity and cost.
- Encouraging innovation and collaboration. Regulators should consider allowing responsible organizations to freely innovate. Organizations that meet criteria for this should be allowed to try new and emerging technologies, with the input and observation of regulators, without any added risk.
- Clarifying, for any given industry, what data can be shared, under what circumstances, and how.
- Clarifying how and when PETs can be used in the context of both cross-industry privacy laws (e.g., CCPA) as well as industry-specific ones (e.g., HIPAA, Gramm-Leach Bliley, etc.), and explaining how to leverage PETs in a legally-compliant manner. For example, data de-identification is often used in the context of HIPAA. This comes with major drawbacks, including an inability to link data sets as well as a reduction in accuracy and precision of results. If it were clearly understood which PETs could be used in here in a regulatory compliant manner, the speed and accuracy of analyses would be significantly improved, which in turn would positively impact patient outcomes.

Third, it is important to understand that collaboration comes at a cost - for an organization to engage in collaboration and use PETs, they may have to train / retrain resources, change business processes, purchase and implement technology, and integrate technologies with legacy systems. To this end, regulators and lawmakers must understand that without laws which simply encourage collaboration (rather than mandate it), will yield less participation than desired. One can again look to the USA PATRIOT Act as an example. Whereas Section 314(a) is mandatory and has full industry participation, Section 314(b) is not. As such, it is often the case that participating institutions devote less resources to 314(b) information sharing requests, meaning that requesting institutions may not receive timely responses. One reason for this is because 314(b) is not mandated, and participation is not incentivized in any way (e.g., by aiding in supervisory exam scores, which today only look at technical compliance to the law). In summary, regulators and lawmakers should consider both regulatory mandates and safe harbors, as well as incentives, to encourage the use of PETs in any given industry.

5. Specific laws that could be used, modified, or introduced to advance PETs

See #4 above.

6. Specific mechanisms, not covered above, that could be used, modified, or introduced to advance PETs

In addition to the legal and regulatory considerations discussed previously, other supporting mechanisms could include:

- Subsidizing the testing and advancement of PETs in priority areas in both the public and private sector, similar to what the Singaporean Government does when it identifies technologies that it would like the private sector to leverage
- Simplifying government procurement processes for priority technologies in priority areas, and/or creating testing "sandbox" environments for both the public and private sectors to utilize, similar to what the ICO does in the United Kingdom
- Challenge prizes (similar to the US/UK PETs Challenge Prize)
- Techsprints (similar to the recent Anti-Corruption Tech Sprint, where Department of State and Department of Treasury were involved)

7. Risks related to PETs adoption

Some of the main risks related to PET adoption are that today's stakeholders may not know how to adequately evaluate PETs for security, interoperability, and performance. Given that these are relatively new technologies tackling the most sensitive of data and analytics, using them without the adequate evaluation could introduce significant risks, and is exactly why open-sourced, standards-compliant security and interoperability, as well as a well-understood evaluation methodology, are so important.

As related to FHE, there are open-sourced encryption libraries like PALISADE and it's second generation version, OpenFHE. PALISADE supports the major publicly known and accepted encryption schemes, and has been publicly released and available for inspection for several years. It has wide use and adoption in the public and private sector, as well as academia. This is indicative of trust in the encryption technology. A key enabler of this adoption is that all the source code is available for third party security inspections. Other PETs and some other homomorphic encryption providers do not open source their technology, meaning it is difficult to understand and verify their technological and security claims and capabilities. This creates risk, as related to the adoption of PETs.

Connected to this topic is that of security and interoperability standards. The leading standards body in this space is HomomorphicEncryption.org, which was co-founded by IBM, Intel, Microsoft, and Duality, and focuses on security and interoperability of Homomorphic Encryption. Standards written by this group have been leveraged by the ISO to create their draft Homomorphic Encryption standards. Not all PETs have such standards, which increases the risks and costs associated with them.

Finally, there has not historically been a generally accepted approach to evaluate and benchmark PETs. To this end, the HEBench Organization was recently founded by Duality, Intel, Deloitte, IBM, Microsoft, and others to assist in this matter by addressing the lack of structured and consistent measurements of full stack performance. This again helps market stakeholders better understand the PETs they are evaluating, which reduces risk.

8. Existing best practices that are helpful for PETs adoption

Many of the best practices which are helpful for PETs adoption all revolve around standardization. For example, some best practices helpful for PETs are:

- Making code and algorithms publicly available for cryptanalysis.
- Identifying and formalizing various explicit security threat models across multiple PETs.
- Developing with standardized quantifiable (mathematically based) security models and levels of security across PETs.
- Mutual understanding and agreement of relevant regulations and/or legal basis to collaborate on data, and
- Explicit regulatory government approval and even participation (e.g., in Estonia, regulators have actively been involved in efforts to facilitate info sharing to fight money laundering, which has yielded positive results).

9. Existing barriers, not covered above, to PETs adoption:

The main barriers have been covered above.

10. Other information that is relevant to the adoption of PETs

The key points relevant to the adoption of PETs has been covered above.

References and Extra Material

[1] Baron, Dr. Joshua. Cooperative Secure Learning (CSL). DARPA, https://www.darpa.mil/program/cooperative-secure-learning.

[2] DARPA Selects Researchers to Accelerate Use of Fully Homomorphic Encryption. DARPA, March 2021, https://www.darpa.mil/news-events/2021-03-08.

[3] Rocher, L., Hendrickx, J. M., & De Montjoye, Y. A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. Nature communications, 10(1), 1-9.

[4] Sweeney, L. (2013). Matching known patients to health records in Washington State data. arXiv preprint arXiv:1307.1370.

[5] Sweeney, L., Abu, A., & Winn, J. (2013). Identifying participants in the personal genome project by name (a re-identification experiment). arXiv preprint arXiv:1304.7605.

[6] Sweeney, L. (2000). Simple demographics often identify people uniquely. Health (San Francisco), 671(2000), 1-34.

[7] 2018 reform of EU data protection rules. European Commission. May 2018. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

[8] Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017, May). Membership inference attacks against machine learning models. In 2017 IEEE symposium on security and privacy (SP) (pp. 3-18). IEEE.

[9] Mouchet, C., Troncoso-Pastoriza, J., Bossuat, J. P., & Hubaux, J. P. (2020). Multiparty homomorphic encryption from ring-learning-with-errors. Cryptology ePrint Archive.

[10] Li, B., Micciancio, D., Schultz, M., & Sorrell, J. (2022). Securing Approximate Homomorphic Encryption Using Differential Privacy. Cryptology ePrint Archive.

[11] SGAxe: How SGX Fails in Practice

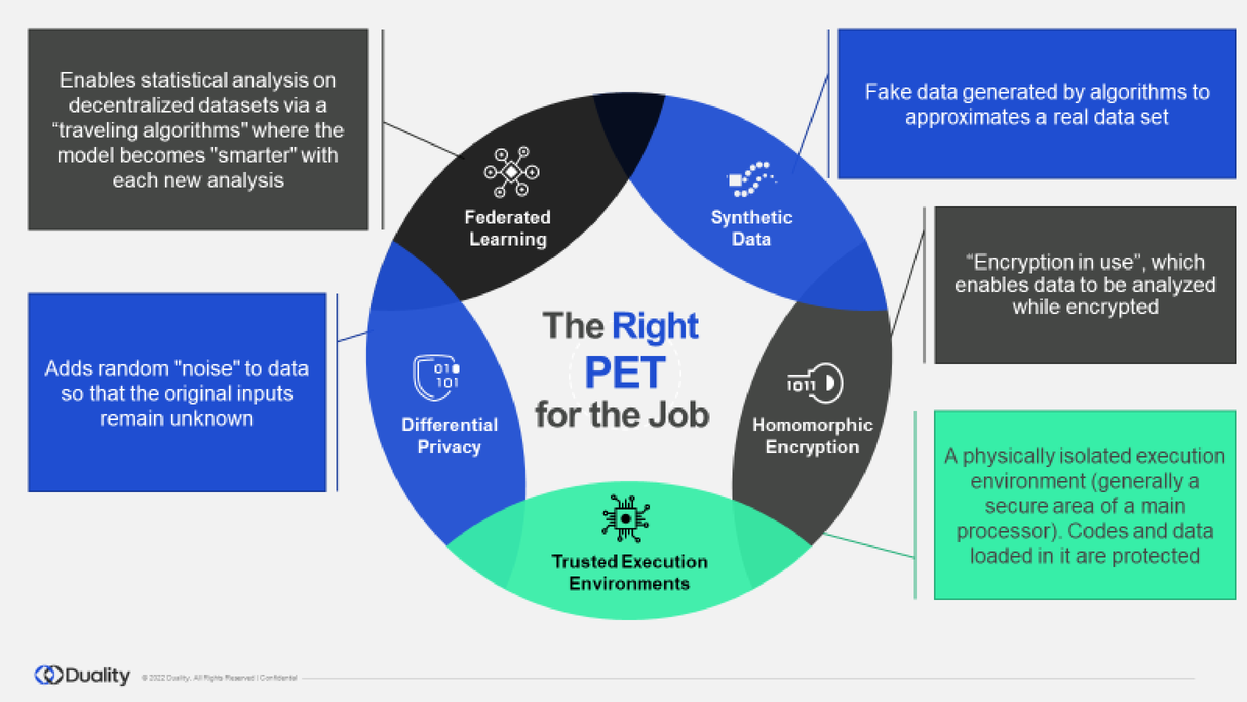
van Schaik, Stephan, Andrew Kwong, Daniel Genkin, and Yuval Yarom. "SGAxe: How SGX fails in practice." 2020. https://sgaxe. com/files/SGAxe. (2020).

[12] Improving 5G Network Security. DARPA, February 2020, https://www.darpa.mil/news-events/2020-02-05.

[13] Section 314(b) Fact Sheet. Financial Crimes Enforcement Network, Dec. 2020, https://www.fincen.gov/sites/default/files/shared/314bfactsheet.pdf.

Appendices

Appendix A – Privacy Enhancing Technologies: Definitions



Enables statistical analysis on decentralized datasets via a “traveling algorithms" where the model becomes "smarter" with each new analysis

Fake data generated by algorithms to approximates a real data set

“Encryption in use", which enables data to be analyzed while encrypted

Adds random "noise" to data so that the original inputs remain unknown

A physically isolated execution environment (generally a secure area of a main processor). Codes and data loaded in it are protected

Duality © 2022 Duality. All Rights Reserved | Confidential