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

Archer, David W.; Varia, Mayank; Smart, Nigel; Malozemoff, Alex; Darais, David; Baum, Carsten; Rosulek, Mike; Tromer, Eran; and Near, Joe

A Response to the Request for Information on Advancing Privacy-Enhancing Technologies by the Office of Science and Technology Policy

David W. Archer, PhD - Principal Scientist, Galois, Inc.
Mayank Varia - Associate Professor, Boston University
Nigel Smart - Professor, KU Leuven; Chief Academic Officer at Zama
Alex Malozemoff, PhD - Principal Researcher, Galois, Inc.
David Darais, PhD - Principal Scientist, Galois, Inc.
Carsten Baum - Assistant Professor, Aarhus University
MIke Rosulek - Associate Professor, Oregon State University
Eran Tromer - Associate Research Scientist, Columbia University
Joe Near - Assistant Professor, University of Vermont

Respondent type: Industry and Academic Collaboration

08 July 2022

Preface

In this response, we address selected questions from the OSTP RFI where we have deep expertise and long experience. Our response team regularly collaborates on PET technology research and development, addressing social and legal implications of PETs, and providing practical proofs of PET usefulness in multiple disciplines. Our viewpoints combine both industry and academia across the USA and northern Europe.

In addition to our direct responses below, we encourage OSTP to also carefully study the work of the United Nations Privacy Preserving Techniques Task Team, accessible at this website: https://unstats.un.org/bigdata/task-teams/privacy/. Members of our response team, particularly Dr. Archer and Prof. Varia, led development of the original UN Handbook of Privacy-Preserving Computation Techniques, initially published in 2018, and are contributing significantly to its second version due later this year. Prof. Varia is also writing a companion guide for the law profession to assess the extent to which PETs satisfy and open new affordances under privacy and data protection regulations.

1. Specific research opportunities to advance PETs:

While PETs have come a long way in the past 10 years, some significant advances must be made to bring them into practical mainstream use, for example in the Government sphere. The remaining areas that we believe need the most attention are:

- **Programmability.** Today, most PET applications are hand-coded and optimized by a small handful of expert cryptographers. This approach is analogous to the machine-code level programming used in the early days of computers, prior to the advent of high-level programming languages (HLLs) and the abstractions they provide. The few attempts at such HLLs for PETs to date have been proofs-of-concept, unsuitable to address practical problems. HLLs and their compiler toolchains for PET programming that are easy to use, yet provide abstractions useful across multiple PET families, are an area of research critical to transitioning PETs into everyday use. We note that the IARPA-funded HECTOR program, unfortunately canceled after its first year, was aimed in significant part at this concern.
- **Debuggability.** Identifying and mitigating defects during the software development process remains a critical productivity drain for all programmers. This problem is dramatically amplified for PET programming, but remains largely unsurfaced because most PET programs so far remain very simple and thus easy to reason about. While typical debugging of PET programs may fall into the realm of “just an engineering problem”, there is a deeper problem: growing insistence by organizations (especially national security activities) that formal methods in computer science be used to mathematically verify programs before adoption. Thus a salient, unaddressed area of research in PETs is the development of formal verification strategies and technical approaches for their verification.

- **Compilation tools.** Today, we rely on purpose-built compilers to transform application source code into executable code, with a distinct compiler that provides specific treatment for each PET framework. For example, compilers for linear secret sharing (LSSS) frameworks aim to minimize the number of communication rounds, while FHE compilers aim to minimize (sometimes similar, but often different) costs such as computation depth and number of bootstrappings. Compilers that successfully accommodate diverse frameworks, yet provide commonality of programmer support, error reporting, and other features is an important area for development.
- **Certification.** New cryptographic approaches are subject to certification by diverse federal authorities (for example, NSA's CyberSecurity Directorate). Approval by those authorities is often much more rapid and successful when the candidate protocols are based on well-understood, already-approved cryptographic primitives. However, nascent protocols for PETs are often developed in academia using novel primitives that do not meet that standard. Thus research in PETs that rely on standard cryptographic methods would benefit the nation in terms of getting PETs to a place where they can be readily certified or authorized to operate in mission settings.
- **PET hardware acceleration in low-power regimes.** PETs may find significant adoption at the network edge, even though current commercial focus often seems to be "in the cloud". Examples of edge processing include tactical computers at forward operating bases (FOBs) processing diverse signal sources that must be kept private; or satellite or ground-station based computation on imagery data that is sensitive, as a precursor to transmitting analytic results to mission users. Because PETs are still generally very resource-intensive, and because computational power at the network edge is always at a premium, special-purpose accelerator hardware specifically designed for low-power operation will be a critical technology to drive adoption. Research is needed on how to design such accelerators to be most efficient in both power and speed, as well as on how to divide the workload of secure computation between software on a typical host CPU and hardware acceleration. We point to the DARPA DPRIVE program as a first effort in this direction.

Today, certain PETs are showing more promise than others. We believe that the most promising PET technologies today are the following:

- Private set intersection (PSI) with (generally statistics-based) computation on the resulting intersection. Statistical analysis across sensitive data held by multiple Government activities or agencies is a growing need, called for in multiple legislative initiatives as well as the Federal Data Strategy. However, the conundrum of how to share such data while respecting the need for its confidentiality continues to confound agencies. PSI offers the most efficient sensitive data linkage approach available, while providing cryptographic protections over the data being shared. We point for example to a recent pilot project at the National Center for Educational Statistics, conducted as a collaboration between Galois, Inc. and Georgetown University, and using PSI technology from Galois to operate at full scale for nationwide data.
- Use of multi-party computation, and in particular linear secret sharing, for example to secure sensitive cryptographic artifacts such as signing keys for digital signature protocol

(including post-quantum signature schemes) is also a growing need with promising PET support. At present, the NIST candidate post-quantum signature schemes such as Crystals/Dilithium do not have such *threshold signing* capability. However, promising prototypes are under construction now, and early indications suggest that performance may be on par with mission-scale throughput.

- Interactive zero knowledge proof protocols for complex proofs such as software cybersecurity look particularly promising. Current leading edge work demonstrates the ability to prove the existence of vulnerabilities such as the OpenSSL Heartbleed weakness, and upcoming enhancements will offer proof of key properties of software such as *memory safety*. Non-interactive ZK proofs that verify correct computation while keeping the inputs to and results of the computation private is another area where ZK PETs are promising, with some relatively-simple instances already deployed in financial-technology applications.
- Fully homomorphic encryption has entered early deployment in a few places, for well-structured applications. As called out above, programmability and debuggability still must be addressed, but FHE (especially with hardware acceleration as needed) appears to be on track to address well-chosen applications.

2. Specific technical aspects or limitations of PETs:

We highlight two technical limitations of PETs:

- First, PETs necessarily and purposefully inhibit the ability to perform manual cleaning and vetting of sensitive input data, because those inputs are by nature encrypted, and because secure computation algorithms for cleaning data are too complex to be computationally efficient. This problem necessitates changes in the data cycle workflow when using PETs. That said, we emphasize that two common techniques remain viable: performing data cleaning at the source *prior* to encryption and subsequent inclusion within a PET-enabled workflow; and performing automated techniques such as outlier detection and data reliability testing with PET protections applied to these techniques along with the rest of the analysis. The latter leverages the expressive power and programmability of these PETs.
- Second, PET statistical disclosure techniques such as *differential privacy* provide strong privacy benefits by preventing computation outputs from revealing input data, at the expense of adding distortion to those outputs. Here, we wish to emphasize that the relaxation of perfect accuracy is inherent to the task of providing output privacy, rather than being a specific flaw of differential privacy. Other attempts at disclosure limitation based on suppression or the use of quasi-identifiers have consistently been shown to be re-identifiable; recent examples include https://arxiv.org/pdf/2202.13470.pdf and https://queue.acm.org/detail.cfm?ref=rss&id=3295691. Additionally, privacy budgets for differential privacy techniques that might appear to be excessive due to their worst-case nature have increasingly been found to have matching attack bounds that are viable to execute in practice.

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

The classes of applications that particularly benefit from PETs are significantly influenced by the performance limitations of PETs today. In particular, (1) the analysis used in the application should occur on a recurring but infrequent basis; (2) The computation itself should be rather straight-forward (e.g., simple statistics vs. training a neural net), if all the data were actually present & centralized. The challenge should not be the computation but rather the inability to collect all relevant data centrally; and (3) participants should be willing to accept some delay in receiving a response. While there do exist high performance and real-time applications for PETs, the reality is that most potential PET applications today would over-stress current data processing systems. However, we point to several applications that meet the criteria above and are thus practical today:

- **Government statistics reporting across multiple organizations**. As mentioned above, OSTP might refer to a report on using PETs in a prototype for the US Department of Education here: https://mccourt.georgetown.edu/news/a-federal-government-privacy-preserving-technology-demonstration/
- **De-confliction of computer network resource use in law enforcement operations**. As described above, *private set intersection* is a particularly promising PET. Securely de-conflicting the use of network vantage points (that is, computers) from which diverse law enforcement agencies monitor and mitigate cyber-crime is a natural fit for such technology.
- **High-security digital signatures**, particularly for national security use. Highly sensitive digital signature keys or other encryption keys should never be materialized in one place. Secure multi-party computation is currently being explored as a practical way to materialize such keys in a secure form that is immune to exfiltration.
- **Cross-border financial crime detection**. The finance industry is often an early adopter of new technology, and emerging statutes such as GDPR make cross-border data sharing impossible. PETs appear to be a natural fit for solving this conundrum. Significant work is already being done in this space.
- **Distinguishing perpetrators from informants**. At the recent Theory and Practice of Multi-Party Computation conference, the firm Roseman Labs provided a nice example: Law enforcement personnel are interested in finding perpetrators of human trafficking, but distinguishing them from known trafficking victims who are in contact with an NGO. A form of private set intersection was used to combine lists of suspects and such victims, across multiple organizations.
- **Enabling pathways to justice for survivors of sexual assault and harrassment**. Survivors of sexual assault and harrassment are often very hesitant to come forward and report such incidents. At the same time, perpetrators are well known for exploiting legal channels to suppress claims of assault and abuse. Research has found that equities are dramatically improved, and survivors are more willing to come forward and take action if they know that others have been victimized by the same perpetrator. Keeping survivor

data private and immune to inappropriate perpetrator suppression is a natural fit for PETs. We encourage OSTP to refer to the work at https://www.mycallisto.org/ . In particular, we refer the reader to this paper: https://www.projectcallisto.org/callisto-cryptographic-approach.pdf.
Dr. Archer and Prof. Varia worked on development of the PETs used there, and served on the cryptography advisory board for that organization.

- **Privacy-preserving digital assets**. Financial privacy is crucial to emerging digital finance systems, such as Central Bank Digital Currencies (CBCD). Indeed, the Federal Reserve opined that "Protecting consumer privacy is critical. Any CBDC would need to strike an appropriate balance, however, between safeguarding the privacy rights of consumers and affording the transparency necessary to deter criminal activity." https://www.federalreserve.gov/publications/files/money-and-payments-20220120.pdf
Zero-knowledge proof systems offer a powerful approach for achieving these aims in blockchain-based systems, as explored by Zerocash (http://zerocash-project.org) and subsequent operational deployments and enhancements.
- **Robustness to supply-chain threats.** Modern IT systems rely on the integrity of their computational components, which may be compromised by supply chain compromise, intrusion, or faults. Zero-knowledge proofs offer a way to assure the integrity of the final output, even if underlying computer or network components are compromised — while protecting the confidentiality of sensitive information implicated in the derivation of that data. Current realizations still entail a high overhead, but show great promise for use in high-assurance critical applications.
- **Data integrity.** A related goal is assuring the integrity of *data* which may have been manipulated in undesired ways. For example, ensuring the truthfulness of images is a major challenge, especially given the evolution of Deep Fake capabilities. Academic technology-demonstrators show that zero-knowledge proofs can assist in vetting authenticity of image files (https://www.cs.tau.ac.il/~tromer/photoproof), and are ripe for extension to other media types and concrete applications.

We also encourage OSTP to engage with the Department of Homeland Security and obtain the presentations given at the recent PETS4HSE (PETs for the Homeland Security Enterprise) workshop there. Dr. Archer's presentation provided several examples of practical PET use suitable for DHS.

6. Specific mechanisms, not covered above, that could be used to advance PETs:

We emphasize strongly that the way to put PETs into practice is to demonstrate what they can do – what new capabilities are achievable with PETs that were not achievable before. History in the realm of cyber security shows clearly that bringing security techniques to bear on current solutions - "sprinkling security over the top" - does not lead to new adoption. Indeed, if a capability is in use today, even insecurely, users are far more likely to accept the current risks and continue, rather than change to a secure mechanism. (This is an unfortunate state of

affairs, because attack patterns always improve over time, making current mechanisms progressively less secure unless protected with cryptographically provable security. We refer for example to the Census Bureau's own example attack on their own previous de-identification technology). Thus what's important for PET adoption is to *prove out new ideas*, focusing on those that are firmly grounded in the needs of practical, real-world use. To that end, we recommend the use of grants to the States to develop PET-based novel solutions that benefit their citizens and that come with firm demonstration metrics for privacy. Statistical applications, such as for example analysis of how life-long educational outcomes correlate to economic wellbeing, may be a good place to start.

9. Barriers, not covered above, to PET adoption:

As with any technology that depends on data sharing, a key barrier to adoption is the *willingness* to share data, whether secure or not. That is, PETs can come into their own only when two or more parties agree to pool data to do something new. We recommend that OSTP review the Lunar Ventures report on the potential business upside of such sharing. The problem is that such coming together to even discuss sharing is hard for companies, or due to anti-trust regulations may be deemed illegal, and in any case struggles against the corporate mindset of protecting intellectual property. Getting over this initial resistance for enough commercial use cases is a salient barrier to wide-spread adoption of PETs commercially. Similar barriers appear in the Government sphere.